The impact of structuring peer feedback in a wiki-based CSCL environment on performance and feedback content

Mario Gielen

Promotor: Prof. dr. Bram De Wever

Proefschrift ingediend tot het behalen van de academische graad van
Doctor in de Pedagogische Wetenschappen

2016

VOORWOORD

Het doel van dit proefschrift is meer inzicht te verwerven in de manier waarop peer assessmentpraktijken geoptimaliseerd kunnen worden, en meer in het bijzonder hoe het structureren van de rol van de beoordeelde en de beoordelaar in het peer feedbackproces een impact kan hebben op de kwaliteit van de prestatie en de peer feedback van lerenden. Net als het onmogelijk is om een allesomvattend antwoord te bieden op alle prangende vragen in de assessment en scripting literatuur, is het onmogelijk alle personen te bedanken die op één of andere manier een invloed hebben gehad op de uitwerking van dit proefschrift. Een aantal van hen wens ik hier in het bijzonder te bedanken.

In de allereerste plaats dank ik mijn promotor, prof. Dr. Bram De Wever, voor de geboden kans om dit proefschrift te schrijven, alsook voor de aandacht die eraan werd geschonken de afgelopen vier jaar. Je deur stond gewoonweg altijd open voor mij en ik heb dan ook ongelooflijk veel van je geleerd. Ook wil ik je bedanken voor het nodige geduld en begrip toen ik ons departement vroegtijdig verliet om aan de slag te kunnen gaan als onderwijskundig adviseur aan UHasselt. Het werd moeilijker en moeilijker om 's avonds geconcentreerd verder te schrijven, maar jouw ondersteuning zorgde ervoor dat ik steeds de bomen door het bos bleef zien.

Verder wens ik de leden van de begeleidingscommissie – prof. dr. Hilde Van Keer, prof. dr. Yves Rosseel en prof. dr. Keith Topping – te bedanken voor de aandacht, tijd en suggesties die ze hebben gespendeerd aan dit proefschrift tijdens de verschillende bijeenkomsten. Ook wil ik prof. dr. Martin Valcke bedanken dat hij zonder aarzelen het praktisch gedeelte van het vak 'Onderwijskunde' vanaf de start van mijn doctoraatstraject aan mij heeft toevertrouwd.

De collega's van de vakgroep hebben via hun collegialiteit, ondersteuning en vriendschap elk op hun eigen manier bijgedragen aan dit proefschrift. Een speciale dank gaat uit naar de mensen van het tecolab onderzoeksteam, de vaste congres compagnons en de plakkers tot in de late uurtjes. Ik heb superleuke momenten beleefd in Gent, Istanbul, Chicago, Munchen, Hong Kong, Tallinn, Madrid, Porto, Cyprus enz. In het bijzonder wil ik mijn lieve bureaugenoten Lisa Dewulf, Hanne Tack en Elise Burny bedanken voor de gezellige en leerzame babbels.

Daarnaast wil ik de studenten van het eerste jaar bedanken om deel te nemen aan mijn tijdsrovende schrijf -en evaluatieopdrachten in de wikis. Ook had ik niet zonder de hulp van Liesje De Backer gekunnen bij het verbeterwerk van duizenden abstracten. Verder wil ik graag Marlou Mespreuve en Donna Willaert bedanken voor hun enthousiasme om in functie van hun masterthesis de kwalitatieve component van mijn onderzoek onder handen te nemen. Ik vond het spijtig dat ik jullie niet tot het einde heb kunnen begeleiden maar ik ben trots op jullie werk. Daarnaast wens ik ook Joke De Lentdecker en Dagmar De Potter te bedanken bij het nodige codeer -en inscanwerk. Ook wil ik de coolste germaniste van België, Lien Cox, in het bijzonder bedanken voor op de meest onmogelijke momenten toch mijn teksten vakkundig en bereidwillig na te lezen en te verbeteren.

Ook wil ik mijn uitgestrekte Limburgse entourage van Koersel en omstreken, de oudleiding van de Chiro, de Homies, de Vrijbuiters, mijn café voetbalploeg, de bende van de Plas en de Kibkes gang allemaal bedanken om me aan alles behalve het werk te laten denken wanneer dat kon. Als enig kind heb ik altijd enorm veel waarde gehecht aan vrienden en vriendinnen die me gevormd hebben en nog steeds vormen tot de persoon die ik ben en die ik wil zijn. Ik ben dan ook heel gelukkig dat ik nu, na een aantal leerrijke wendingen in binnen –en buitenland, terug bij mijn Limburgse roots terechtgekomen ben. Ook nog een uitzonderlijk woord van dank aan Lesley Williamson, die gedurende de eerste 3 jaar van mijn doctoraatstraject als vriendin betrokken was bij mijn dagdagelijkse wereldschokkende ontdekkingen of frustaties rond 'peer assessment'.

Tot slot verdienen mijn ouders een bijzondere plaats in dit voorwoord. Jullie hebben me zien groeien, ook doorheen dit doctoraatstraject. Zonder jullie was dit simpelweg allemaal niet mogelijk geweest. Bedankt voor jullie onvoorwaardelijke steun en kansen die jullie me geboden hebben in mijn leven. Jullie gedrevenheid en doorzettingsvermogen blijven me nog steeds inspireren en verbazen. Mama, bedankt om er altijd voor me te zijn en om de dingen te leren relativeren. Papa, bedankt voor je onvoorwaardelijke interesse waar ik mee bezig was en om af en toe mee te reizen naar de andere kant van de wereld. Ook hielp jullie nuchtere blik me eraan denken dat de wereld niet zou vergaan met of zonder al die A1's. Ik hou van jullie met heel mijn hart.

Lieve Ellen, ik ben blij dat we elkaar al zo lang kennen zodat je maar al te goed wist dat ik ook een ander leven heb buiten het schrijven van dit proefschrift en het ontwikkelen van een MOOC. Ik wil je bedanken voor jouw begrip en steun tijdens deze laatste '*doe-door-maanden'*. Jij weet wanneer me ruimte te geven, en wanneer te voorkomen dat ik afdwaal van waar het essentieel om draait. Bedankt voor te zijn wie je bent en rust te brengen in mijn leven. Het schenkt vertrouwen in onze toekomst.

Hasselt, maart 2016

Mario Gielen

TABLE OF CONTENTS

1 General introduction

Chapter 1
General introduction

"In short, to stay true to our educational values, we must seize the opportunity to reimagine what we do and how we do it."

- Leo Rafael Reif (2013) -

Abstract

By means of general introduction, this chapter describes the context in which the subsequent chapters of this dissertation are situated in detail. The first section of the general introduction will outline the general theoretical background: *Formative assessment*, *Peer Assessment (PA) for Learning, Peer Feedback (PFB), and Structuring PA Practices*. The second section describes the research objectives, which attempt to gain insight into how the PFB process in PA practices can be optimized in function of students' learning. The third section discusses the design and outline of the subsequent intervention studies in order to clarify how the different chapters are related to each other.

Introduction

Formative assessment

A large body of research highlights the power of assessment for the learning process (Evans, 2013; García, García-Álvarez, Moreno, 2014; Kennedy, Chan, Fok, & Yu, 2008; Pellegrino, Chudowsky, & Glaser, 2001). While summative assessment is traditionally a process to assess students' learning outcomes, the main goal of formative assessment is to close the gap between current and desired performance (Sadler, 1989). In this respect, formative assessment aims at providing rich feedback and supporting students' learning (Black & William, 1998). As well as formative assessment offers learners an indication of not only their strengths and weaknesses, it also offers them information about the next steps to be taken in the learning process. Therefore, formative approaches are now commonly accepted, as they focus on promoting and enhancing student learning (Sambell, McDowell, & Montgomery, 2013). In addition, these formative approaches are not only focusing on assessment and feedback received from the instructor. However, students, too often, view feedback as the "responsibility of someone else, usually teachers, whose job it is to provide feedback information by deciding for the students how well they are going, what the goals are, and what to do next" (Hattie & Timperley, 2007, p. 101).

Nowadays, research underlines that the students themselves could be the definitive source of formative assessment (Andrade, 2010). For this reason, self and peer assessment are now common practices of formative assessment in function of students' learning (Nicol & Macfarlane-Dick, 2006). Self-assessment is a process of formative assessment, in which students are required to reflect on the quality of their own work, judge whether it corresponds with the explicitly stated goals or criteria, and revise accordingly (Andrade, 2010). As such, the primary purpose of engaging students in self-assessment is to boost learning and achievement. When engaging students through peer assessment (PA), they are required to reflect on the quality of a peer's performance, and discuss how well it corresponds with the explicitly stated goals or criteria as well as suggesting how the quality of the work could be improved during revision accordingly (Strijbos & Sluijsmans, 2010). As this practice of formative assessment, in which learners collaborate and interact with each other, has been highly praised as an important component of a participatory culture of learning (Kollar & Fischer, 2010), we will attempt to shed more light on PA in this dissertation.

Related to this, previous research highlighted that social interaction such as giving and receiving help is essential for collaboration (Kreijns, Kirschner, & Jochems, 2003). During the past decades, research has been mainly focusing on several forms of collaborative learning, in which learners attempt to achieve educational goals through a coordinated and shared activity, existed already in different ways and forms (Siampou, Komis, & Tselios, 2014). In the present PhD dissertation, the focus will be more on peer feedback within collaborative learning in higher education, since providing and receiving peer feedback are essential parts of the PFB process in PA practices.

Peer Assessment *for* Learning

Inspired by the power of assessment for the learning process, PA is perceived as a common practice of formative assessment, in which students are required to be actively involved in all phases of the assessment process (Dysthe, 2004; Boud & Molloy, 2013). Strijbos and Sluijsmans (2010) described PA as "an educational arrangement where students judge a peers performance quantitatively and/or qualitatively and which stimulates students to reflect, discuss and collaborate" (p. 265). In other words, it is a process whereby peers take on the role of assessor or/and assessee to reflect on "the amount, level, value, worth, quality or success of the product or outcomes of learning of peers" (Topping, 1998, p. 250). As an important component of a participatory culture of learning (Kollar & Fischer, 2010), PA is often suggested as a good approach to increase students' engagement with their own learning (e.g. Nicol & MacFarlane-Dick, 2006; De Wever, et al., 2011). When correcting the work of other group members and providing feedback in small groups, PA appeared to be advantageous both as a learning tool (Topping, 1998) and as an assessment tool (Cheng & Warren, 1997), for the assessor (e.g. Evans, 2013) as well as the assessee (e.g. Tsivitanidou, Zacharia, & Hovardas, 2011). As a learning tool, PA involves learners directly in the learning process and provides them with skills to assess

criteria that define high-quality work (Topping, 1998). As an assessment tool, PA increases the responsibility towards students (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010) as they are actively involved in their own assessment and consequently more engaged in their own learning (De Wever, et al., 2011).

Nowadays, PA research and practices are not only focusing on PA as a "mark-giving" activity. In addition to offering marks, it is crucial for PA practices that peers provide feedback as well. Previous research pointed out that providing and receiving feedback is often perceived as an educational activity to enhance students' learning (eg. Falchikov, 1995), which has the potential to augment students' performance (eg. Falchikov, 2003). In view of PA practices, it is rather logical that students should be given the opportunity to use the received PFB, in order to improve their learning and achievement (Nicol & MacFarlane-Dick, 2004). Therefore, providing and receiving PFB can be seen as the learning element of PA (Liu & Carless (2006). In comparison with giving marks, not that much research can be found on these particular PA components (Evans, 2013). For this reason, the present PhD dissertation will specifically focus on providing and receiving PFB, in which peers juggle with "information provided by an external agent regarding some aspect(s) of the learner's task performance, intended to modify the learner's cognition, motivation and/or behaviour" (Duijnhouwer, Prins, & Stokking, 2012, p. 171).

However, students require a various set of skills in order to act as peer assessor, or to handle feedback from a peer as an assessee. More specifically, the assessor needs to be capable to deal with specific assessment criteria to judge a peer's performance. When students are expected to provide marks, for example scoring rubrics appeared to be particularly helpful because they present the assessment criteria in a structured format (Panadero, Romero, & Strijbos, 2013) and they give an indication on expected performance by listing the relevant assessment criteria and by defining the quality levels of each criterion (Andrade & Valtcheva, 2009). To become skilled peer assessors and assessees, who provide and receive high quality PFB, students require practice and training (Sluijsmans, et al., 2002; Birenbaum, 1996). As such, research stresses that the design of PA activities should be well-thought and incorporate practice occasions in function of students' learning (Strijbos & Sluijsmans, 2010).

Given the rapid development of computer technology in educational settings (Yang & Tsai, 2010), previous research has been focusing on online peer assessment (e.g., Cho & Schunn, 2007). Recently, research emphasized the educational value of online PA for students' learning (Cheng, Liang, & Tsai, 2015). Previous research stated that online assessment is more beneficial compared to face-to-face assessment (Tsai 2009; Tsai & Liang, 2009; Yang & Tsai, 2010). A particular reason could be that online technology offers learners more freedom in time and space (Tsai, Lin, & Yuan, 2002). In this respect, a growing body of research illustrated that peers' performance can be enhanced, when learners are involved learners in online PA practices (e.g. Cheng, Liang, & Tsai, 2015; Xiao & Lucking, 2008). For this reason, the focus of this dissertation will be on online PA, in which students are required to provide and receive PFB in an online environment. However, other research stresses that the social-affective component, as well as

the skills for interactive process management are more difficult to develop in an online environment (McLuckie & Topping, 2004). Related to this, previous research claims that lacking social cues, such as non-verbal behaviour or actual presence, can have a negative impact on online learning (Tu & McIsaac, 2002).

Peer Feedback

In view of formative assessment, the goal is that students should be given the opportunity to use the feedback they receive, in order to improve their learning and achievement (Nicol & MacFarlane-Dick, 2004). Students require feedback that informs them not only if they dealt with particular criteria correctly or not, but also why and what they should do about it to improve their work (e.g. Coll, Rochera, & De Gispert, 2014). For this reason, feedback is seen as an essential factor for progress (Carless, 2015). As feedback demonstrated to have a powerful impact on both learning and performance, it is important to identify the specific feedback features that should be included or excluded, bearing in mind that some types of feedback appeared to be more effective than others (e.g. Nelson & Schunn, 2008). In order to be effective, feedback should tackle three major feedback questions: 'Where am I going?', 'How am I going?', and 'Where to next?' (Hattie and Timperley, 2007). In order to compose a valuable feedback message, the assessor is required to formulate feedback, in which he or she indicates how well the performance matches the expected performance, as well as feed forward, in which he or she suggests how future performance can be improved (Hattie & Timperley, 2007). A feedback message should also comprise suggestions for future improvement, as feed forward component in a feedback message activates the learner in function of future performance (Carless, 2007). More particular, research found that specific and elaborated feedback leads to improved performance and outcomes (Strijbos, Narciss, & Dünnebier, 2010).

When examining prior research on PA, we can notice that within the field of PA practices, especially providing and receiving PFB are perceived as important educational components of the PFB process in PA practices (e.g. Falchikov, 1995). However, we cannot assume that all students will be competent to offer high quality feedback for several reasons such as proficiency (e.g. Kaufmann & Schunn, 2010; Strijbos, Narciss, & Dünnebier, 2010). As students are engaged in high-level cognitive processing during the PFB process (King, 2002), they require skills such as summarising, explaining, providing feedback and identifying mistakes and gaps (Van Lehn, et al., 1995). However, we cannot neglect that not all learners will be able to offer high-quality feedback, due to limited competence, lack of objectivity or insufficient knowledge of how to interact with others (Cheng, Liang, and Tsai, 2015; Strijbos, Narciss, & Dünnebier, 2010). Related to this, Topping (1998) states that the quality of PFB is not comparable to the quality of instructor feedback, but "its immediacy, frequency, and volume compensate for this'' (p. 255). For this reason, students traditionally need to "critically review the PFB they have received, decide which changes are necessary in order to improve their work and to proceed making those changes" (for detailed description see, Hovardas, Tsivitanidou, & Zacharia, 2014, p. 135).

As not all students will be proficient peer assessors from the start, more research is necessary to find out how students can be supported in the PFB process order to safeguard high quality PFB (Hovardas, et al., 2014).

Previously, research has illustrated that PFB could be beneficial for students' learning (e.g. Topping, 1998), such as enhancing the quality of the students' writing (Van Zundert, Sluijsmans, & Van Merriënboer, 2010). During the PFB process, students need to "engage with and take ownership of evaluation criteria, to make informed judgments about the quality of the work of others, to formulate and articulate these judgments in written form and, fundamentally, they need to possess the ability to evaluate and improve one's own work based on these processes"(Nicol, Thomson, & Breslin, 2014, p. 120). Based on the review study of Evans (2013), we can conclude that results on the impact of feedback are rather unpredictable and therefore, more research on the *impact of PFB on learning and performance is needed* (e.g. Hattie and Timperley, 2007).

Feedback quality

Related to this, a large body of research claims that the effectiveness of a feedback message largely depends on its content, template, and function (Narciss, 2006, 2008; Narciss & Huth, 2004; Shute, 2008). In order to examine the quality of PFB messages, Prins, Sluijsmans, and Kirschner (2006) developed the Feedback Quality Index (FQI), which is a scoring rubric that gives an indication whether the PFB message meets certain criteria such as offering explanations, questions, suggestions, examples, etc. In chapter 3 and 5, we incorporate the FQI to determine the quality of the PFB messages. However, previous research emphasised that the specific feedback content appears to be crucial for the impact of feedback on learning and performance (Cho & MacArthur, 2010). In order to answer the question: what determines the actual quality of feedback content? Research emphasised that an effective feedback message should comprise both verifications and elaborations (e.g., Narciss, 2008). In order to examine the specific content of the feedback in more detail, the need of an effective instrument arises to examine the actual composition of these PFB messages in more detail, which on its turn can be a valuable theoretical contribution for the PFB literature. For this reason, a supplementary goal of this dissertation is to develop a content analysis scheme to examine the specific content of PFB messages, which students provide to each other during writing and assessment activities in an online learning environment.

Inspired by the feedback framework of Narciss (2008), Strijbos, Van Goozen, and Prins (2012) have previously developed a coding scheme to analyse PFB messages. Based on this, we further elaborated on this model and added categories, which particularly investigate the style, type and focus of these messages to analyse PFB messages in detail. As such, we can check whether PFB messages contain rather positive or more negative verifications, which tell us whether certain criteria are met or not. Or whether they contain rather informative and

suggestive elaborations that offer relevant information for the learner when revising (Hattie & Gan, 2011). In a later phase, we add two additional categories to examine the level of agreement and implementation when students are requested to evaluate the PFB they receive from the assessor. In chapter 4 and 6, we elaborate on these specific categories of the content analysis scheme.

Structuring Peer Assessment Practices

In the literature, it becomes apparent that so far many questions remain unanswered on how formative assessment should be implemented into educational practice to boost students' learning in higher education (Sadler, 2010). More particularly, research in the PA field has underlined that a rigid approach is lacking on how PA practices, in which students are required to provide and receive PFB, should be customized in function of students' learning (Strijbos & Sluijsmans, 2010). With this dissertation, we aim to shed more light on how this PFB process should be tailored, when students are engaged in writing and assessment activities in the first year of higher education.

As PA is an example of a complex learning task, previous studies have expressed a need of instructional support when students are being involved in high-level cognitive processing such as critical thinking, problem solving, and decision making (e.g. Weinberger, Stegmann, Fischer, & Mandl, 2007; Cole, 2009). Social interaction such as giving and receiving are essential aspects of both PA and collaborative learning. As research underlines that these high-level processes are not happening spontaneously (Kollar & Fischer, 2010). And in order to "trigger engagement in social and cognitive activities that would otherwise occur rarely or not at all" (Kobbe, et al., 2007, p. 212), collaboration scripts are being recommended to successfully improve these types of collaborative learning, in which the activities of all actors are specified, scheduled and delegated during collaborative learning activities (Fischer, Kollar, Stegmann, & Wecker, 2013). In this respect, research underlines that it is challenging to come up with a definition of 'collaborative learning' and 'collaboration', due to its wide variety of uses of the term inside each academic field (Dillenbourg, 1999). In this dissertation, students primarily collaborate with each other by providing and receiving PFB comments in function of improving their own and unique writing performance, instead of collaborating in order to achieve one common writing performance. More specifically, students had to participate in online writing assignments and PA activities, in which they were engaged in a research-based task that was requiring them to read a research paper, to write a concise abstract in a wiki environment, and to provide feedback on other students' abstracts.

However, it appears that more research is necessary, examining the required type of structure and support that peer assessors need in order to produce high quality feedback messages (Hovardas, et al., 2014). Additionally, other research has emphasised the importance of examining the effectiveness of different scaffolding and scripting techniques for PA as an

important direction for future research (Kollar & Fischer, 2010). In this manner, empirical studies examining different kinds of support can be found in the literature (e.g. Quintana, et al., 2004; Kollar, Fischer & Slotta, 2007). Since providing structure in the PA process appears to be a crucial factor to generate high quality feedback (Poverjuc, Brook, & Wray, 2012), a growing body of research has experimented in varying instructional interventions in the PA process to enhance the effectiveness of PFB, for instance, by organizing a training to improve PFB (Sluijsmans et al., 2002), or offering guiding questions to support the assessor while formulating PFB (Gielen & De Wever, 2012), or providing sentence openers to encourage interaction between students (Baker & Lund, 1996). It should be stressed that the existing knowledge is based upon instructional interventions to optimize the PA process; these are primarily studies with a main emphasis on the role of the assessor (Gielen, et al., 2010). As it is important to actively engage all actors in the PFB process, it becomes clear that there is a gap in the literature, which describes best practices of instructional interventions during the PA activities, not only with respect to the role of the assessor, but also and especially in function of the role of the assessee, which has often been neglected so far.

Unfortunately, most research and practices regarding written feedback in higher education appears to focus primarily on a one-way interaction (Nicol, 2010), in which the assessor merely formulates a feedback 'monologue' about the assessees' performance. For this reason, research highlights that written feedback should be conceptualised as a dialogical and contingent two-way process, in which the active engagement of students is mandatory for a co-ordinated interaction between peers (e.g. Kollar & Fischer, 2010; Carless, 2015). The idea that dialogue is fundamental in successful learning and teaching is well established in the educational literature (Nicol, 2010). Therefore, the purpose of dialogue is to help students understand concepts and ideas and to apply their understanding in learning tasks (Laurillard, 2002). Previously, research has highlighted that dialogic feedback processes should be adaptive, discursive, interactive, and reflective (Laurillard, 2013). In order to optimize the PFB dialogue, the process should be *adaptive*, and therefore it is essential that the PFB depends on students' needs. As previous research recommends encouraging the assessee to formulate on which particular components feedback is needed (Nicol & MacFarlane-Dick, 2006; Gibbs & Simpson, 2004). Following, the PFB dialogue should be *interactive* and therefore, linked to actions related to a task goal. Before actually formulating feedback, assessors will need to profoundly process a peers' work and prepare the feedback accordingly (Kollar & Fischer, 2010). For dialogue to be characterised as *discursive*, it needs to exist of a rich two-way communication between assessor and assessee. As dialogue needs to be *reflective* as well, students should be given the opportunity to reflect on the received feedback. In this view, it seems to be essential that assessees evaluate their received PFB, by implementing 'a posteriori reply form' (Gielen et al., 2010) or 'back-feedback' (Kim, 2009), in order to give an indication if the assessor's feedback is for example relevant, helpful, and implementable. In view of closing the feedback loop (Boud, 2000), previous research stressed the importance of taking into account the response of the assessee in a final phase of the feedback cycle (Gibbs & Simpson, 2004), which generates a so-called feedback dialogue (Prins & Mainhard, 2009). In requesting feedback, students engage in reflection even before

comments are received, while in responding to feedback they are interactive in linking their feedback comments to the assignment task (Nicol, 2010). For this reason, we involve the assessee purposely more actively at the start and at the end of the PFB process during the last intervention of this dissertation.

All things considered, the overall purpose of the feedback process in higher education is to help students develop the ability to monitor, evaluate and regulate their own learning (Nicol & Macfarlane-Dick, 2006). Inspired by this, this dissertation attempts to find answers *on how the roles of both assessor and assessee in PA practices should be tailored in order to optimize a dialogic PFB process in function of students' learning and performance*. To achieve this main aim in the next section, we will first elaborate on the three research objectives and secondly on the design and outline of the subsequent intervention studies in order to clarify how the different chapters of this dissertation are related to one another.

Research design and overview of the dissertation

Research objectives

Building on the shortcomings mentioned earlier in the research context of this chapter, the main aim of this dissertation is to gain insight into how the PFB process in PA practices can be optimized in function of students' learning. More specifically, the aim of this dissertation is to examine how to structure the role of the assessor and assessee in order to optimize the PFB process so that it can have an impact on the quality of students' performance during PA practices (RO1) and PFB (RO2). This general aim is divided into two general research objectives that directed the different studies of this dissertation. As a third research objective, which is a sub goal of RO2, the aim is to analyse the actual content of the provided PFB messages in more detail (RO3).

Research objective 1 (RO1): To explore the impact of different levels of structuring in the PFB process on students' performance

Research objective 2 (RO2): To explore the impact of different levels of structuring in the PFB process on the quality of students' PFB messages, measured by the Feedback Quality Index (Prins, Sluijsmans & Kirschner, 2006).

Research objective 3 (RO3): To develop a content analysis scheme, which can be used to examine the specific content of students' PFB messages into more detail; and to implement this content analysis scheme to explore the impact of different levels of structuring in the PFB process on the content of students' PFB messages.

Design and outline of the studies

This dissertation entails seven chapters wherein, besides an introductory (Chapter 1) and concluding chapter (Chapter 7), five chapters are included (Chapters 2 to 6) that report on the results of three interventions in 5 studies with a quasi-experimental design. The data has been previously collected in a wiki-based CSCL environment in the first year of higher education, in which students were expected to participate in online writing assignments and PA activities. Students were engaged in a research-based task that was requiring them to read a research paper, to write a concise abstract in the wiki environment, and to provide feedback on other

students' abstracts. The three succeeding interventions have two main foci. Dealing with RO1 and RO2, the first focus of this dissertation emphasizes on identifying possible discrepancies in students' writing performance and PFB quality, measured with scoring rubrics, when assessor and/or assessee receive a different level of structuring in the PFB process. Regarding RO1 and RO2, the output of study 1, 2 and 4 is respectively described in Chapter 2, 3 and 5. In order to shed more light on the actual PFB content quality, the second focus of this dissertation, RO3 dealt with the development of a content analysis scheme with the main purpose to examine the specific content of PFB message, when students provide feedback to each other during writing assignment in a CSCL environment in higher education. Regarding RO3, the output of study 3 and 5 I respectively reported in Chapter 4 and Chapter 6, as illustrated in Figure 1. At the end of this introductory chapter, Table 1 offers an overview of the different chapters, in which more details of the particular focus, research design and sample of the three intervention studies are given, combined with their analysis methods in function of the three research objectives. Following, the outline of the seven chapters will be discussed.

Chapter 1 provides a **general introduction** to this dissertation. First, it describes the research background of assessment in higher education. Secondly, an overview is given of the different studies and chapters integrated in this dissertation. Lastly, the design and outline of the subsequent intervention studies are discussed in order to clarify how the different chapters relate to one another.

In chapter 2 entitled *'Peer assessment in a wiki: Product improvement, students' learning and perception regarding PFB'*, the added value of offering structure in the PA process is examined in **study 1**. The aim of this chapter is to reveal differences in (1) learning effect, (2) wiki product improvement and (3) students' perception towards the PA process in a CSCL environment in higher education. Study 1 involves two conditions: structured PFB (S-PFB) and non-structured (control). At this point, no other requirements are specified for the role of assessor and assessee, as the focus is to study the impact of the provided structure in the PFB template. Prior to this data were collected from 179 students, who were divided into 38 groups of maximum 5 students. For this purpose, ANOVA's were used to identify differences in students' learning effect and product improvement, when structure is provided in the PA process. Also, Independent-samples t-tests were employed to identify differences in students' perception towards PA. This chapter is published in the *Procedia of Social and Behavioral Sciences*.

In **study 2 (Chapter 3)**, the primary research goal is to explore the impact of a PFB template with a different structuring degree for assessor in the PFB process during PA practices. The present study involves three conditions: a no structure, a basic structure and an elaborate structure condition. To this aim, a quasi-experimental study was set up at the end of 2012 with 168 first-year bachelor students divided into 37 groups in an online wiki environment. For this assignment, which was kept identical in study 4, students had to write three times a draft and final version of an abstract of a submitted, yet not published scientific article related to the topic (i.e., they received the paper, but the abstract was omitted). Before writing the final version, they

received PFB on their draft version. In function of RO1 and RO2, the impact of different levels of structuring in the PFB process on students' writing performance and PFB quality and discussed in Chapter 3 '*Structuring the Peer Assessment Process: A Multilevel Approach for the Impact on Product Improvement and PFB Quality'.*

This chapter is based on an article that was published in 2015 in the *Journal of Computer-Assisted Learning*. In order to tackle RO3, we are particularly interested in finding out more about the specific feedback content students actually provide to each other. For this reason, it was necessary to develop a content analysis scheme, which can be used to examine the specific content of students' PFB messages during writing assignments in a CSCL environment in higher education.

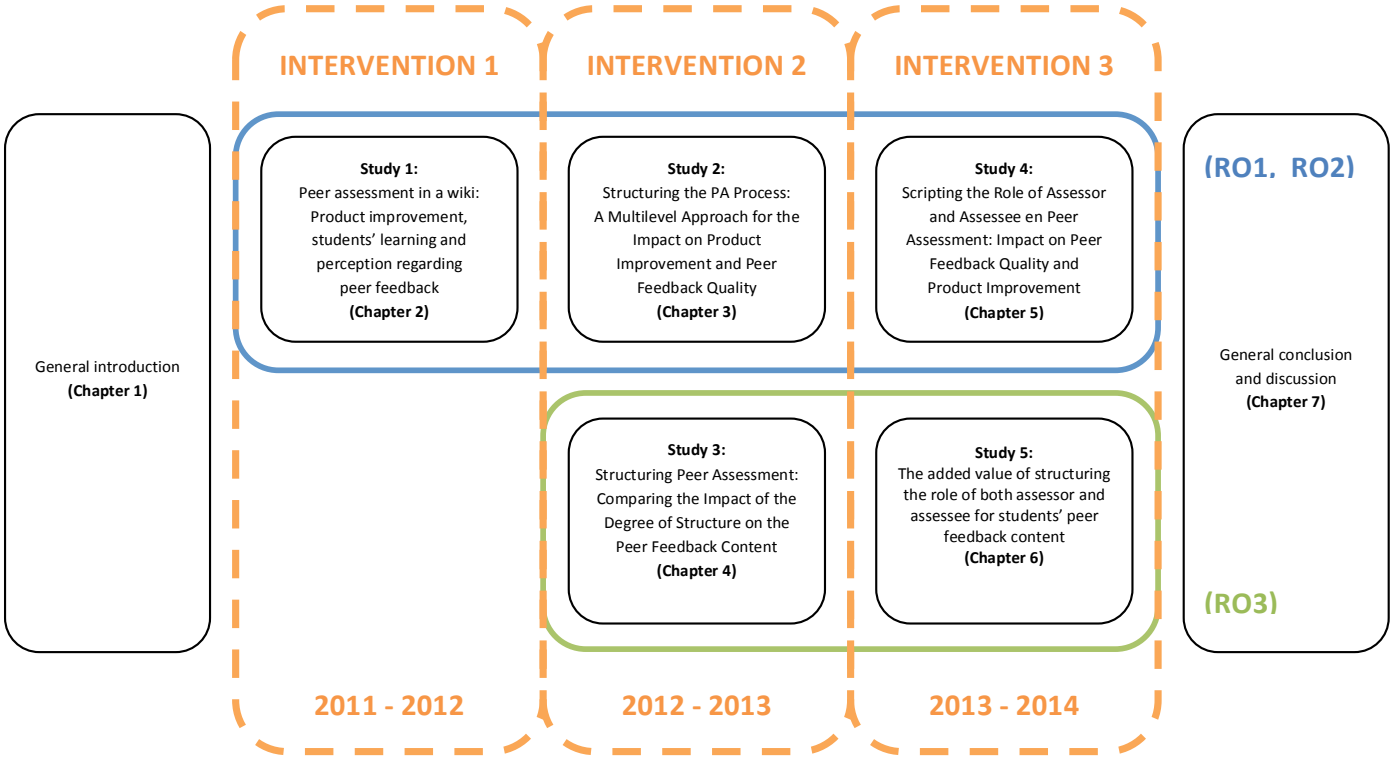


Figure 1. Overview of the 3 different interventions and 5 studies

Study 3 (Chapter 4) '*Structuring Peer Assessment: Comparing the Impact of the Degree of Structure on the PFB Content*' examines the impact on the PFB content quality in more detail, when assessors receive a PFB template with a different level of structuring. By performing a quantitative content analysis of the PFB messages of 41 students, and analyses of (co)variance, this chapter attempts to detect possible differences between the conditions regarding the proportion of PFB content categories: (1) PFB style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus. Chapter 4 is based on an article that was published in 2015 in *Computers in Human Behavior*.

Next, **study 4 (Chapter 5)** examines the added value of an instructional intervention, when both the assessor and assessee are actively involved in the PFB process. More particular, the aim of study is to examine the effect of structuring the role of the assessee and/or assessor by respectively providing them with a PFB request and/or content checklist, together with a structured PFB template. This study adopts a 2x2 factorial design, in which four conditions are compared: (1) a control condition, (2) a feedback request condition, (3) a content checklist condition, and (4) a combination (feedback request + content checklist) condition. Chapter 5 '*Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on PFB quality and product improvement*' examines the impact of engaging both the assessor and assessee in the PFB process, on the quality of students' PFB and written product in a wiki-based CSCL environment in the first year of higher education. To this aim, a quasi-experimental study was set up at the end of 2013 with 125 first-year bachelor students in an online wiki environment. Multilevel analysis was performed to examine the effect of time, student and group level influences on students' PFB quality and product scores. This chapter is based on an article that was published in 2015 in *Computers and Education*.

In function of RO3, **study 5 (Chapter 6)** '*Structuring Peer Assessment: Comparing the Impact of the Degree of Structure on PFB Content*' examines whether further structuring the role of the assessee (asking and evaluating feedback) and the assessor (preparing and providing feedback), on top of providing PFB with the help of a structured PFB template, can have an impact on the PFB content. Multilevel analysis is performed to indicate differences between the conditions regarding the occurrence of the following PFB content categories: (1) PFB style, (2) verification type, (3) verification focus, (4) elaboration type, (5) elaboration focus, (6) evaluation agreement, and finally (7) evaluation implementation. The results of this chapter are presented in an article accepted for publication in the *Journal of Computer-Assisted Learning*.

Finally, chapter 7 contains the **general discussion** on the presented studies in the dissertation related to the main proposed research questions. Further, also strengths, limitations and future research aspirations are proposed. This chapter concludes with contributions and implications for research, practice, and policy.

	Chapter	Focus of the intervention studies	Research design and sample	Data-analysis techniques	RO
	CH1	General introduction (theoretical framework, research questions, research design and overview of the dissertation)			
INT.1	**CH2**	Study 1 investigates the added value of offering structure in the PA process. The aim of this chapter was to reveal differences as to (a) the learning effect, (b) the wiki product improvement and (c) students' perceptions of the PA process.	Quasi-experimental design *Sample:* Groups (n=38) Students (n=179)	Repeated measures ANOVA (SPSS) Repeated measures ANOVA (SPSS) Independent-samples t-tests (SPSS)	RO1
INTERVENTION2	**CH3**	Study 2 examines the impact of different levels of structuring in a PFB template for the assessor on (a) students' writing performance and (b) PFB quality.	Quasi-experimental design *Sample:* Groups (n=37) Students (n=168) Measurement occasions (n=3)	Multilevel analysis (MLwiN)	RO1 RO2
	CH4	Study 3 takes a closer look at the specific peer feedback content, when different levels of structuring are implemented in a PFB template in the PFB process of the assessor, regarding the occurrence of PFB content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus.	Quasi-experimental design *Sample:* Groups (n=9) Students (n=41) PFB forms (n=123) PFB segments (n=4717)	Content analysis AN(C)OVA (SPSS) MAN(C)OVA (SPSS)	RO3
INTERVENTION3	**CH5**	Study 4 investigates how an instructional intervention focused on engaging both the assessor (content checklist) and the assessee (PFB request) in the PFB process can have an impact on (a) students' writing performance (a) and (b) PFB quality.	Quasi-experimental design 2x2 factorial design *Sample:* Groups (n=27) Students (n=125) Measurement occasions (n=3)	Multilevel analysis (MLwiN)	RO1 RO2
	CH6	Study 5 takes a closer look at the specific peer feedback content, when both the assessor (content checklist) and the assessee (PFB request) in the PFB process receive additional structure over time, on top of providing and evaluating PFB with the help of a structured PFB template, regarding the occurrence of the following PFB content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, (5) elaboration focus, (6) evaluation agreement, and finally (7) evaluation implementation	Quasi-experimental design 2x2 factorial design. *Sample:* Groups (n=16) Students (n=79) PFB forms (n=237) PFB segments (n=8440)	Content analysis Multilevel analysis (MLwiN)	RO3
	CH7	General conclusion and discussion (overview and discussion of the main results, limitations and suggestions for future research, and implications)			

Table 1. Overview of three different interventions throughout the dissertation

References

Andrade, H. L. (2010). Students as the definitive source of formative assessment. *Handbook of formative assessment*, 90-105.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice, 48*, 12–19.

Baker, M. J., & Lund, K. (1996). Flexibly structuring the interaction in a CSCL environment. In P. Brna, A. Paiva & J. A. Self (Eds.), *European conference on artificial intelligence in education* (pp. 401–407). Lisbon, Portugal.

Birenbaum, M. (1996). Assessment 2000: towards a pluralistic approach to assessment. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). Boston, MA: Kluwer.

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

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151–167.

Boud, D., & Molloy, E. (2013). Rethinking Models of Feedback for Learning: The Challenge of Design. *Assessment & Evaluation in Higher Education, 38*, 698-712.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*, 57-66.

Carless, D. (2015). *Excellence in University Assessment: Learning from Award-winning Practice*. London: Routledge.

Cheng, K.H., Liang J.C., & Tsai, C.C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78-84.

Cheng, W., & Warren, M. (1997). Having second thoughts: Students perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239.

Cho, K., & MacArthur, C. (2010) Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Cho, K. & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline. *Computers and Education, 48*, 409–426.

Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers and Education, 52*, 141–146.

Coll, C., Rochera, M.J., & De Gispert, I. (2014). Supporting online collaborative learning in small groups: teacher feedback on learning content, academic task and social participation. *Computers & Education, 75*, 53-64.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14*, 201-206.

Dillenbourg P. (1999) What do you mean by collaborative leraning?. In P. Dillenbourg (Ed) *Collaborative-learning: Cognitive and Computational Approaches*. (pp.1-19). Oxford: Elsevier

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction, 22*, 171–184.

Dysthe, O. (2004). *The challenges of assessment in a new learning culture*. The 32nd International NERA/NFPF Conference, Reykjavik, Iceland.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83*, 70-120.

Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 157-166). London: Kogan Page.

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching, 3*, 102–108.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66.

García, A. S., García-Álvarez, M. T., & Moreno, B. (2014). Analysis of assessment opportunities of learning spaces: On-line versus face-to-face methodologies. *Computers in Human Behavior, 30*, 372-377.

Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning, *Learning and Teaching in Higher Education, 1*, 3–31.

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.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585-594.

Hattie, J. & Gan, M. (2011). Instruction based on feedback. In Mayer, R. E. & Alexander, P. (eds). *Handbook of research on learning and instruction* (pp. 249-271). NewYork: Routledge. Taylor and Francis Group.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of educational research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152.

Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*, 387-406.

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.

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment and Evaluation in Higher Education, 34*, 105–114.

King, A. (2002). Structuring Peer Interaction to Promote High-Level Cognitive Processing. *Theory Into Practice, 41*, 33-39.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., Hakkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 211-224.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708-721.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335-353.

Laurillard, D. (2002). Rethinking teaching for the knowledge society. EDUCAUSE review, 37(1), 16-24.

Laurillard, D. (2013). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. Routledge.

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

McLuckie, J., & Topping, K. J. (2004). *Transferable skills for online peer learning. Assessment & Evaluation in Higher Education, 29*, 563-584.

Narciss, S. (2006). *Informatives Tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse (Informative Tutorial Feedback)*. Münster, Waxmann.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner & R. Brunken (Eds.) *Instructional design for multimedia learning* (pp. 181-195). Münster, Waxmann.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*, 501–517.

Nicol, D.J. and Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice. In C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. Ross & B. Smith (Eds.) *Enhancing Student Learning Through Effective Formative Feedback* (pp. 3-14). York: The Higher Education Academy.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*, 102-122.

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*, 195-203.

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

Prins, F., & Mainhard, M. T. (2009, August). *Fostering student's self-regulation during feedback dialogues in vocational education.* Paper presented at the 13th biennial conference of the European Association for Research on Learning and Instruction, Amsterdam, The Netherlands.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465-477.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., ... & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences, 13*, 337-386.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–44.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*, 535-550.

Sambell, K., McDowell, L., & Montgomery C. (2013). *Assessment for Learning in Higher Education*. London: Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Siampou, F., Komis, V., & Tselios, N. (2014). Online versus face-to-face collaboration in the context of a computer-supported modeling task. *Computers in Human Behavior, 37*, 369–376.

Sluijsmans, D., Brand-Gruwel, S., Van Merriënboer, J. J. G. & Bastiaens, T. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*, 23–42.

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.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages*. Paper presented at the EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

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

Tsai, C. C. (2009). Internet-based peer assessment in high school settings. In L. T. W. Hin, & R. Subramaniam (Eds.), *Handbook of research on new media literacy at the K-12 level: Issues and challenges* (pp. 743–754). Hershey, PA: Information Science Reference.

Tsai, C. C., & Liang, J. C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science, 37*, 293–310.

Tsai, C. C., Lin, S. S. J., and Yuan, S. M. (2002). 'Developing science activities through a networked peer assessment system'. *Computers & Education 38*, 241–252.

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506-519.

Tu, C. H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American journal of distance education, 16*, 131-150.

Van Lehn, K.A., Chi, M. T., Baggett, W. , & Murray, R. C. (1995). *Progress report: Towards a theory of learning during tutoring*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburg.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270-279.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge—cognitive, computational and educational perspectives* (pp. 191–211). New York: Springer.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*, 186–193.

Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*, 72–83.

2 Peer assessment in a wiki: Product improvement, students' learning and perception regarding peer feedback

This chapter is based on:

Gielen, M., & De Wever, B. (2012). Peer assessment in a wiki: Product improvement, students' learning and perception regarding peer feedback. *Procedia-Social and Behavioral Sciences, 69*, 585-594.

Chapter 2
Peer assessment in a wiki: Product improvement, students' learning and perception regarding peer feedback

Abstract

The present study examines the added value of peer assessment in a computer-supported collaborative learning environment (CSCL) in higher education by focusing on (1) the learning effect, (2) wiki product improvement and (3) students' perception of peer feedback in a CSCL-environment. The present study involved two conditions: structured peer feedback (S-PFB) and non-structured (control). The results do not indicate a significant learning effect between pretest and posttest or between the conditions. However, for both conditions the peer feedback process improved significantly the quality of the wiki product from draft to final version, although no significant differences between the control and the experimental group (S-PFB) were found. Furthermore, the S-PFB group adopted a more critical attitude when providing and receiving peer feedback. The S-PFB group also perceived the received peer feedback as being more profound and detailed.

Introduction

Peer assessment (PA) has been highly praised as an important component of a participatory culture of learning (Kollar & Fischer, 2010). Therefore, peer assessment might be an instructional strategy with the potential to correspond to the so-called imperatives of twenty-first century pedagogy: customization, interaction and learner-control (Collins & Halverson, 2009). To increase the potential impact of peer assessment on learning, it is crucial "to understand which mechanisms affect learning, and how these mechanisms can be supported" (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010, p. 304). This introduction will start off by focusing on how peer assessment can be associated with 'assessment for learning'. After this, we take a closer look at wikis as computer-supported learning (CSCL) environments to facilitate collaborative learning and peer assessment. Finally, we discuss students' learning effect, product improvement, and perception regarding (structuring) peer assessment in a CSCL-environment.

Peer assessment for learning

The traditional perception of learning has shifted towards a more participatory culture of learning where learners collaborate and interact with each other. Therefore, modern education aims at self-directed and collaborative learning (Boud, Cohen, & Sampson, 1999). These new approaches of learning and instruction require new assessment practices (Strijbos & Sluijsmans, 2010). According to previous studies, this shift from '*assessment of learning*' towards '*assessment for learning*' requires students to become active participants in all phases of the assessment process (Dysthe, 2004). More specifically, assessment gives learners an indication of their strengths and weaknesses but also of the next steps to be taken in the learning process. In this respect, formative assessment aims at providing rich feedback and supporting learning (Black & William, 1998). The main goal of formative assessment is to close the gap between current and desired performance (Sadler, 1989). Therefore, feedback can be perceived as a practice of formative assessment to improve, accelerate, and self-regulate learning. Several studies highlight the power of assessment on the learning process (Black & Wiliam, 1998; Kennedy, Chan, Fok, & Yu, 2008; Pellegrino, Chudowsky, & Glaser, 2001). As peer assessment is a common practice of formative assessment, it is "an educational arrangement where students judge a peers performance quantitatively and/or qualitatively and which stimulates students to reflect, discuss and collaborate" (Strijbos & Sluijsmans, 2010, p. 265). In other words, it is a process whereby peers take on the role of assessor or/and assessee (van Zundert, Sluijsmans, & van Merriënboer, 2010) to reflect on "the amount, level, value, worth, quality or success of the product or outcomes of learning of peers" (Topping, 1998, p. 250). Previous research has shown that peer assessment has a positive influence on the learning process both as a learning tool (Topping, 1998) and as an assessment tool (Cheng & Warren, 2000). As a learning tool, peer assessment involves learners directly in the learning process and provides them with skills to assess criteria that define high-quality work (Topping, 1998). As an assessment tool, peer assessment increases the responsibility towards students (Gielen, Dochy, & Onghena, 2010) as they are actively involved into their own assessment and consequently more engaged in their own learning (De Wever et al., 2011).

Wikis as a tool for CSCL

Collaborative learning is a "mutual engagement of participants in a coordinated effort to solve the problem together" (Rochelle & Teasley, 1995, p. 70). Its secret to success is social interaction, such as giving and receiving help (Kreijns, Kirschner, & Jochems, 2003; Liaw & Huang, 2000; Northrup, 2001). Constructivism is perceived as the underlying learning theory for computer-supported collaborative learning (CSCL) (Kirschner, Martens, & Strijbos, 2004). CSCL is a learning approach, where learners collaborate on authentic problems and issues in an educational online environment (Jacobson & Wilensky, 2006). Collaborative learning systems are designed "to concentrate on refining, integrating, and facilitating the learning process and content knowledge of students during collaborative activities" (Kumar, Gress, Hadwin, & Winne, 2010, p. 826). In contrast with traditional environments, CSCL-environments are promising to merge learners' present state with the intended learning outcomes (Hattie & Timperley, 2007), by offering rich

educational experiences as a preparation for students (Reich, Murnane, & Willett, 2012). Previous research has proved that CSCL-environments have the potential to support and evaluate the regulation process (Soller, Martinez Monés, Jermann, & Muehlenbrock, 2005), the discovery learning process (De Jong, 2006), or the communication process (Saab, et al., 2007). A CSCL-environment simulates classroom situations by "providing shared work- spaces, on-line presentations, lecture notes, reference material, quizzes, student evaluation scores, and facilities for chat and online discussions" (Kumar, et al., 2010, p. 826). Regarding the communication aspect, research emphasises on the importance in CSCL research to take into account features that trigger students' motivation to actively participate in online discussions (Naranjo, Onrubia, & Segués, 2012). As a CSCL-environment, a wiki can be perceived as an interesting tool for individual or collaborative content creation (De Wever, et al., 2011). More specifically, wikis are an interesting learning environment for group assignments to work, write, share and construct knowledge together with other peers (Elgort, Smith, & Toland, 2008). Previous research found that students use wikis for a great diversity of learning activities, such as "to publish homework assignments, maintain portfolios, peer review writing, post artwork, download music for rehearsals, and review drills for physical education" (Reich, et al., 2012, p.10). As each contribution of every student in a wiki is published online, wikis have great potential for facilitating peer assessment (Xiao & Lucking 2008). Therefore, Kollar and Fischer (2009) highlight that peer assessment is an important feature to take into account when educators design learning environments.

(Structuring) Peer Feedback: Effects on learning, product, perception and attitudes

A self-fulfilling prophecy can also be discerned regarding different science class types or tracks (Eder, 1981). In many educational systems, including the Flemish system, students are separated into different academic tracks, which consist of a package of courses focusing on languages, economics, and/or science, respectively. Dividing students into different academic tracks occurs primarily on the basis of their proficiency as determined by previous course grades, yet tracking also occurs based on student and parent choice (Pickens & Eick, 2009). Studies have shown that differences in teacher expectations regarding student science learning, caused by tracking, also result in a differing quality of teaching science (Nieswandt & Shanahan, 2008; Pickens & Eick, 2009). Students in science tracks receive instruction emphasizing scientific reasoning and inquiry-based instruction (Haury & Milbourne, 1999), whereas general-track students receive less challenging instruction, and are subsequently less motivated to learn science (Oakes, 2005).

Peer feedback is perceived as an approach of peer assessment, peer feedback aims to involve students in assessment for learning by students giving each other opinions, suggestions and ideas. Kaufman & Schunn (2010) stress the need for peer assessment research across settings and subjects to find out more on how students' perception and attitudes affect their performance.

Hence, students' perception and attitudes towards peer assessment in CSCL will be examined in more detail in this study. Previous research discovered that students sometimes perceive peer assessment as unfair and often question peers' qualifications to review and assess their work (Kaufmann & Schunn, 2010; Strijbos, Narciss, & Dünnebier, 2010). Additionally, Topping (1998) states that the quality of peer feedback is not comparable with the quality of instructor feedback, but "its immediacy, frequency, and volume compensate for this'' (p. 255). Kaufmann and Schunn (2010) summarize the following strategies to improve students' perception towards peer assessment: enlarging students' peer assessment experience (Sluijsmans et al. 2001; Wen & Tsai 2006); clarifying peer assessment criteria (Falchikov 2005; Smith, Cooper, & Lancaster, 2002); and providing training and support in the peer assessment process (Cheng & Warren 1997; Falchikov 2005, 2007).

Regarding the quality and quantity of feedback in peer assessment of writing, Strijbos et al. (2010) highlight that more specific and elaborated feedback leads to better performance and outcomes. More specifically, process feedback has an impact on learning, students' satisfaction and functioning of the group, while performance feedback seems to improve performance (Gabelica, Bossche, Segers, & Gijselaers, (2011). Although the learning effects of providing elaborated feedback are relatively obvious, Van der Pol et al. (2008) stated that receiving peer feedback, which depend on the feedback quality and assessor's expertise, does not automatically results in significant learning effects. A review study by Hattie and Timperley (2007) revealed that further research on the impact of peer feedback on learning and achievement is required. Therefore, this first part of this study focuses initially on the learning effect of receiving and providing peer feedback, but also on the actual performance by evaluating the final product of the wiki assignment. In a CSCL-environment, instructors have the opportunity to structure the collaboration assignment and feedback process to a certain extent. Particularly, CSCL has the potential to facilitate students' learning when structure or instructional support is foreseen (Strijbos & Weinberger, 2010; Fischer, Kollar, Mandl, & Haake, 2007; Järvelä, Häkkinen, Arvaja, & Leinonen, 2004; Kirschner & Kreijns, 2005; Schellens & Valcke, 2006; Strijbos, De Laat, Martens, & Jochems, 2005) They can provide scaffolds to support the cognitive processes and also to fairly divide the workload and responsibility between group members (O' Donnell, 1999). Additionally, the peer feedback process encourages students' critical thinking skills (Berg, 1999). Regarding structuring interaction, the literature mostly refers to aspects such as roles, facilitated by scripts and prompts, modelling, and specific task and communication instructions (King, 1999). Also, the aspect of '*anonymity*', offers various advantages in CSCL-environments (Ainsworth, Gelmini-Hornsby, Threapleton, Crook, O'Malley, & Buda, 2011). According to Morris, Church, Hadwin, Gress, & Winne (2010), it would be valuable to examine "the extent to which interaction should be structured on an epistemic level in order to support the way learners cope with the uncertain situation of online learning" (Morris et al., 2010, p. 818). Since Strijbos and Weinberger (2010) underline benefits of offering structure in a CSCL-environment, the last part of this study focuses on students' perception when structuring the peer assessment process.

Research questions

For this study, we have formulated the following research questions:

Question 1: Is there a difference (1a) in product quality before and after the PA process and (1b) between the two conditions?

Question 2: Is there (2a) an increase in learning effect from pretest to posttest and (2b) between the two conditions?

Question 3: Has structuring the PA process an effect on students' perception?

Method

Learning Participants, setting and research design

The participants in the present study were first-year bachelor students Educational Sciences (N = 179), enrolled in the course Instructional Sciences at Ghent University. During the collaborative phase, students could access the wiki anywhere and anytime. As shown in Figure 1, this study adopted a quasi-experimental research design. All students were requested to fill in a questionnaire before and after the group assignment. Students were randomly assigned to groups (N = 38) of maximum five students to collaborate on one wiki. Groups were randomly assigned to a condition: either the control condition (n = 19) or the experimental condition (n = 19). In total, 85 students were in the control condition and 94 were in the experimental condition. . From the start, both groups had access to a general introduction movie on general feed back principles, accessible in the electronic learning environment. The assignments were organized as such that after intermediate peer feedback, students always had the time to revise their draft version into the final text. In the experimental feedback condition, called the structured peer feedback (S-PFB) condition, the instructor offered students a peer feedback template with two additional guiding questions, besides a bullet-pointed criteria list, while the control feedback condition had no specific format to provide peer feedback.

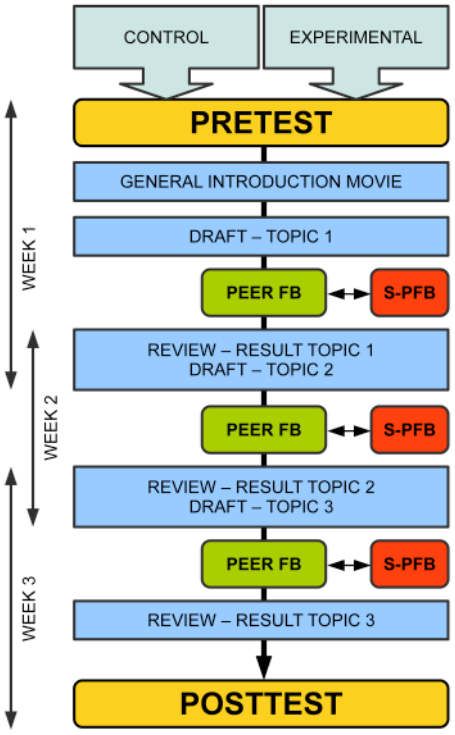


Figure 1 - Research Design

Wiki assignment

This course had a blended design, in which the weekly course lectures were combined with wiki assignments throughout the semester. The total grade of three wiki assignments took up 40% of the final course grade (the other 60% was based on a theoretical written exam). In total, the practical part of the subject consisted out of 4 wiki assignments. The first one was a trial wiki to get familiar with the educational technology and expectations. The duration of each of the 3 other wikis was 3 weeks, so 9 weeks in total. This research was based on the findings of the last wiki assignment. In this final wiki assignment, students had to collaborate in constructing a wiki by tackling previously used exam questions based on theory of the three main topics of the course, which were already taught in the lectures: behaviorism, cognitivism and constructivism, in which students had to receive and provide peer feedback. After receiving peer feedback, the assessee was requested to review the draft version of the previous main topic into the final version. At the end of the three-week period, each group submitted one wiki, including the final work of all five group members, resulting in a wiki consisting of ten questions on the three topics (5 students x 2 questions x 3 topics = wiki with 30 questions). Each question includes a draft version, peer feedback, and a final version. The learning environment provided for each group a wiki area as CSCL-environment.

Research instruments

The pretest, which was completed by the students before the start of the writing assignment, consisted out of six multiple-choice exam questions to examine the insight on the three main topics. Therefore, they had to choose one answer out of four options. After the submission of the wiki assignment, students were requested to fill in a posttest. This posttest consisted out of six comparable multiple-choice exam questions, to examine the insight on the three main topics after the wiki assignment. Once again, they had to choose one answer out of four options.

In addition to the posttest, students were asked to fill out a questionnaire at the end of the wiki-assignment. This questionnaire was divided into four subsections. Variables in this study were the students' preferences, perceptions and attitudes. All items were measured using 5-point Likert scales, and anchored by 1 (totally disagree) and 5 (totally agree). The first section investigated the perception of students towards the writing assignment in a wiki environment (e.g., "I am satisfied about the contribution of the other group members") (14 items). The second and third section examined the perception of students respectively towards receiving feedback (e.g., "I consider the received feedback as relevant") (11 items) and providing feedback (e.g., "I consider my provided feedback as profound and detailed") (13 items). The fourth section evaluated general feedback assumptions (e.g., "I prefer to provide feedback anonymously") (4 items).

Results

Quality of the wiki product

In order to answer the first research question, namely if the quality of the wiki product increases from draft to result (1a), a repeated measures ANOVA found that there was a significant difference between the quality of the initial work and the quality of the final product, $F(1, 175) = 390.399$, $p = <.001$. As shown in Table 1, a repeated measures ANOVA found for the second part of the first research question (1b) that there was no significant difference between the quality of the product between the control and S-PFB condition, $F(1, 175) = 3.533$, $p = .062$.

Table 1

Quality of Product of Wiki Assignment

	Draft		Result	
Comparison of quality of group work	M	SD	M	SD
S-PFB condition	14.47	3.77	19.56	2.32
Control condition	15.89	3.34	20.10	1.75

Note. *** $p<.001$

Learning effect of the feedback process

To answer the second research question, namely if the feedback process has a learning effect on the students between the pretest and the posttest (2a), a repeated measures ANOVA found no significant difference in learning effect between the pretest and posttest scores of the multiple-choice questions, $F(1, 175) = .005$, $p = .945$. As shown in table 2, a repeated measures ANOVA did not show a significant difference between the learning effect between the control and S-PFB condition (2b), $F(1, 175) = .405$, $p = .525$.

Table 2

Learning Effect of Feedback Process

	Pretest MCQ		Posttest MCQ	
Scores on multiple choice questions (MCQ)	M	SD	M	SD
S-PFB condition	2.43	1.41	2.52	1.76
Control condition	2.42	1.18	2.31	1.71

Note. *** $p<.001$

Structuring peer feedback: students' perception and attitudes

In order to answer the third research question, namely if structuring the PA process has an effect on students' perception, independent-samples t-tests were conducted. Table 3 provides an overview of students' perception towards providing and receiving structured peer feedback between the two conditions. Regarding providing peer feedback, students' perception in the experimental condition (M=4.20, SD=.720) claims to be more critical towards giving peer feedback than students in the control condition (M=3.90, SD=.715), who had no structure at all for giving feedback. An independent samples t-test showed this difference to be significant; t(155)=2.584, *p*=.011. Regarding receiving peer feedback, students who made use of a S-PFB form indicated to be more critical towards receiving peer feedback (M=3.81, SD=.702), than students from the control group (M=3.56, SD=.748). An independent samples t-test showed this difference to be significant; t(154)=2.184, *p*=.030.

An independent-samples t-test also revealed a significant difference in students' perception towards how profound and detailed the received peer feedback is. The S-PFB (M=3.16, SD=.833) group perceived the *received* peer feedback as more profound and detailed than the control (M=2.86, SD=.698) group; t(153)=-2.372, p=0.19. On the other hand, there was no significant difference found regarding the profoundness and level of detail of *provided* peer feedback between the experimental (M=3.40, SD=.805) group and the control (M=3.19, SD=.762) group; t(155)=1.633, p=104. Cohen's d statistics indicate a small to moderate effect size for students' perception towards giving and providing peer feedback.

Table 3

Students' Perception towards Providing and Receiving Peer Feedback

	Control (PFB)		Exp. (S-PFB)				
Perception towards Peer Feedback	M	SD	M	SD	*t*	*df*	*d*
A wiki is an ideal tool to facilitate the process of providing and receiving peer feedback	3.30	1.18	3.44	.957	.745	130[i]	-0.13
I consider receiving peer feedback as an added value	4.11	.688	4.08	.732	.256	153	-0.66
I consider providing peer feedback as an added value	3.79	.730	3.87	.799	-.641	155	-0.12
I am critical towards the received peer feedback	3.56	.748	3.81	.702	-2.18*	154	-0.34
I am critical towards the work of a peer when providing feedback	3.90	.715	4.20	.720	-2.58*	155	-0.42
I consider the received peer feedback as profound and detailed	2.86	.698	3.16	.833	-2.37*	153	-0.39
I consider the given peer feedback as profound and detailed	3.19	.762	3.40	.805	-1.63	155	-0.27
I am satisfied with the quality of the received peer feedback	3.63	.759	3.71	.834	-.666	153	-0.10
I am satisfied with the quality of the given peer feedback	3.74	.628	3.81	.627	-.753	155	-0.11
I prefer providing peer feedback anonymously	2.14	1.04	2.16	1.01	-.158	155	-0.02

Note. [i] Equal variances not assumed (Hildebrand et al. 2005: 362); * p <.05

Discussion

The findings in this study showed that the quality of the wiki product clearly improves after receiving peer feedback. This is in agreement with the research of Kaufman and Schunn (2010), who suggest that improvements could be linked to the higher quantity or a more positive nature of peer feedback. In this respect, Nadler (1979) argues by stating that individual-level feedback provided in a team setting improves individual performance. Previous studies, which examined the quality and quantity of peer feedback in writing assignments, revealed that more specific and elaborated feedback stimulates better performance and outcomes (Strijbos, et al., 2010). In this study, the results showed a significant difference between the draft and final version of the wiki product. On the other hand, there is no significant difference in quality of the product between the two conditions.

"Feedback guides, motivates, and reinforces effective behaviors and reduces or stops ineffective behaviors" (London, 2003, p. 1). Although feedback is widely accepted as an appreciated tool for learning (Falchikov, 2001), the findings of the present study did not show a significant learning effect between the pretest and posttest. Previous research of Alvero, et al. (2001) emphasizes that studies are ambiguous regarding requirements for the effectiveness of feedback. This is in agreement with the review study by Hattie and Timperley (2007). The results also showed no difference in learning effect between the experimental and the control condition. As far as we know, there is no research that investigates the difference in learning effect when structuring the peer feedback process. Additionally, previous research pointed out a difference in learning effect between providing and receiving feedback (Van Der Pol, et al., 2008). By providing peer feedback, it is essential that students "invest time and effort into actively constructing content-oriented reactions" (Van der Pol, et al., 2008, p.1816). The learning effect of receiving peer feedback depends on the quality of the feedback by the assessor, who is not an expert. Previous studies highlight that the competence level of the assessor influences students' perception towards feedback (Strijbos, et al., 2010).

Regarding the perception of students towards structuring peer feedback, the results pointed out that students, who gave and received peer feedback with the help of S-PFB, had a more critical attitude in the feedback process, than the other students. This finding is in agreement with Berg (1999), who discovered that peer feedback stimulates critical thinking. The findings also showed that students, who used S-PFB in the feedback process, perceive the received peer feedback as being more profound and detailed. Li, Liu, and Steckelberg (2010) discovered that students acknowledge the value of peer feedback, but that they were not always satisfied about the quality of their received peer feedback. The lack of constructive and more detailed feedback was associated with poor quality feedback. In general, previous research highlights that CSCL facilitate students' learning, especially when the instructor offers some kind of support or structure to concretize the roles and activities of the involved students (Strijbos & Weinberger, 2010; Fischer, et al., 2007; Kirschner & Kreijns, 2005; Schellens & Valcke, 2006; Strijbos, et al., 2005). Therefore, we might assume that offering a little structure to provide peer feedback helps students in a

certain extent through the different steps of the thinking process when they are requested to provide profound and detailed peer feedback.

Limitations and implications for further research

When interpreting the results, the limitations of the study should be taken into account. First of all, it is worth mentioning that the multiple-choice questions, which were used to calculate the learning effect, were not calibrated accurately. The lecturer of Educational Sciences selected the different questions for the pretest and posttest, but the equality of pretest and posttest was not statistically tested. Another limitation is the basic and limited structure of the S-PFB form. The present study revealed several gaps in existing research on structuring peer assessment in CSCL that provide starting points for future experimental research. Further research is needed to investigate the impact of different S-PFB approaches (e.g. a more elaborated structured feedback form) and the influence of students' critical attitude on product improvement and students' learning and perception.

Conclusion

Although the literature suggests several benefits of the peer assessment process, the results of this study do not show a significant difference in learning effect from pretest to posttest, and also not between the two conditions. Based on the received peer feedback, the wiki product improves significantly when students have the opportunity to revise their initial draft before submitting the final result in the wiki-based CSCL-environment. The results point out that the basic intervention of S-PFB through a feedback form does not have a significantly additional impact on students' final product. When structuring the peer assessment process in a wiki-based CSCL-environment, this study revealed that students have a stronger critical attitude when they both provide and receive peer feedback with the help of S-PFB. To conclude, students experience received peer feedback as being more profound and detailed when the feedback is constructed with the help of S-PFB.

Acknowledgements

This study was supported by the Ghent University BOF Research Grant B/11676/02.

References

Ainsworth, S., Gelmini-Hornsby, G., Threapleton, K., Crook, C., O'Malley, C., & Buda, M. (2011). Anonymity in classroom voting and debating. *Learning and Instruction, 21*, 365-378.

Alvero, A., Bucklin, B., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985–1998). *Journal of Organizational Behavior Management, 21*, 3–30.

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing, 8*, 215-241.

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

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education, 24*, 413-426.

Cheng, W., & Warren, M. (1997). Having second thoughts: Students perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239.

Collins, A. & Halverson, R. (2009). *Rethinking Education In The Age Of Technology: The Digital Revolution And Schooling In America*. New York: Teacher College Press.

De Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen & D. Clark (Eds.), *Handling complexity in learning environments: research and theory* (pp. 107–128). London: Elsevier Science Publishers.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14*, 201-206.

Dysthe, O. (2004). *The challenges of assessment in a new learning culture*. The 32nd International NERA/NFPF Conference, Reykjavik, Iceland.

Elgort, I., Smith, A. G., & Toland, J. (2008). Is wiki an effective platform for group course work? *Australian Journal of Educational Technology, 24*, 195–210.

Falchikov, N. (2005). *Improving assessment through student involvement.* New York: Routledge Falmer.

Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 128–143). New York: Routledge. G.

Fischer, F., Kollar, I., Mandl, H., & Haake, J. M. (Eds.). (2007). *Scripting computer- supported collaborative learning: Cognitive, computational and educational perspectives*. New York: Springer.

Gabelica, C., Bossche, P. V. D., Segers, M., & Gijselaers, W. (2011). Feedback, a powerful lever in teams: A review. *Educational Research Review, 7*, 123-144.

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.

Gielen, S., Dochy, F., & Onghena, P. (2010). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*,137–155.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*, 81-112.

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences, 15*, 11–34.

Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*, 387-406.

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.

King, A. (1999). Discourse patterns for mediating peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87–117). Mahwah, NJ: Lawrence Erlbaum Associates.

Kirschner, P. A., & Kreijns, K. (2005). The sociability of computer-mediated collaborative learning environments: Pitfalls of social interaction and how to avoid them. In R. Bromme, F. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication: And how they may be overcome* (pp. 169–192). Dordrecht: Kluwer.

Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in Higher Education? A Framework for Designing Multiple Collaborative Environments. *Computer-Supported Collaborative Learning Series, 3*, 3-30).

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335-353.

Kumar, V. S., Gress, C. L. Z., Hadwin, A. F., & Winne, P. H. (2010). Assessing process in CSCL: An ontological approach. *Computers in Human Behavior, 26*, 825-834.

Liaw, S., & Huang, H. (2000). Enhancing interactivity in web-based instruction: A review of the literature. *Educational Technology, 40*, 41-45.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*, 525-536.

London, M. (2003). *Job feedback: Giving, seeking, and using feedback for performance improvement* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Morris, R., Church, H., Hadwin, A.F., Gress, C.L.Z., & Winne, P.H. (2010). The use of roles, scripts, and prompts to support CSLC in gStudy. *Computers & Human Behavior, 26*, 815-824.

Nadler, D. A. (1979). The effects of feedback on task group behavior: a review of the experimental research. *Organization Behavior and Human Performance, 23*, 309-338.

Naranjo, M., Onrubia, J., & Segués, M. T. (2012). Participation and cognitive quality profiles in an online discussion forum. *British Journal of Educational Technology, 43*, 282-294.

Northrup, P. T. (2001). A framework for designing interactivity into web-based instruction. *Educational Technology, 41*, 31-39.

O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 179–196). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Reich, J., Murnane, R., & Willett, J. (2012). The State of Wiki Usage in U.S. K–12 Schools: Leveraging Web 2.0 Data Warehouses to Assess Quality and Equity in Online Learning Environments. *Educational Researcher, 41*, 7-15.

Rochelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). New York: Springer-Verlag.

Saab, N., Van Joolingen, W. R., & Van Hout-Wolters, B. H. A. M. (2007). Supporting communication in a collaborative discovery learning environment: The effect of instruction. *Instructional Science, 35*, 73-98.

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

Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education, 46*, 349–370.

Sluijsmans, D. M. A., Moerkerke, G., van Merrienboer, J. J. G., & Dochy, F. J. R. C. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation, 27*, 153–173.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*, 71–81.

Soller, A., Martinez Monés, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education, 15*, 261–290.

Strijbos, J. W., De Laat, M. F., Martens, R. L., & Jochems, W. M. G. (2005). Functional versus spontaneous roles during CSCL. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Computer supported collaborative learning 2005: The next 10 years!* (pp. 647-656). Mahwah, NJ: Lawrence Erlbaum Associates.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

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.

Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*, 491-494.

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

van der Pol, J., van den Berg, B. a. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*, 1804-1817.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270-279.

Wen, M. L., & Tsai, C.C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 27*, 27–44.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*, 186-193.

3 Structuring the Peer Assessment Process: A Multilevel Approach for the Impact on Product Improvement and Peer Feedback Quality

This chapter is based on:

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449.

Chapter 3
Structuring the Peer Assessment Process: A Multilevel Approach for the Impact on Product Improvement and Peer Feedback Quality

Abstract

In order to optimize students' peer feedback processes, this study investigates how an instructional intervention in the peer assessment process (PA) can have a beneficial effect on students' performance in a wiki environment in first-year higher education. The main aim was to study the effect of integrating a peer feedback template with a varying structuring degree. The present study involved three conditions: a no structure, a basic structure, and an elaborate structure condition. Due to a clear hierarchical structure, in which over time (level 1), 168 students (level 2) are nested within 37 groups (level 3), multilevel analysis was performed to examine the effect of time, student and group level influences on students' peer feedback quality and product scores. Results revealed that both peer feedback quality and product scores increase significantly for all conditions over time, after multiple practice occasions. In addition, after several practice occasions, significant differences were found between the conditions in both peer feedback (elaborate higher than no structure) and product scores (elaborate and basic higher than no structure). Building on this, limitations, directions for future research, and practical implications are presented.

Introduction

Peer feedback as educational practice

Recently, several authors (eg. Black & William, 1998) have emphasised on the power of assessment for the learning process, rather than assessment of learning. This focus on formative rather than on summative assessment aims at identifying, and ultimately closing, the gap between current and desired performance (Sadler, 1989). Peer assessment (PA) is one specific method that can be employed for formative assessment and thus reach this aim while involving learners in all phases of the assessment process (Dysthe, 2004). Previously, peer assessment is defined as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal- status learners" (Topping, 2009, p. 20). Therefore, PA is often suggested as a good approach for increasing students' engagement with their own learning (eg. Nicol & MacFarlane-Dick, 2006). To this day however, research on PA in higher education

remains "very variable in type and quality, scattered and fragmentary in nature" (Topping, 1998, p. 267; see also Evans, 2013, who presents the same conclusion).

In the context of PA, peer feedback is often perceived as an educational activity for enhancing students' learning (eg. Falchikov, 1995), in which peers juggle with "information provided by an external agent regarding some aspect(s) of the learner's task performance, intended to modify the learner's cognition, motivation and/or behaviour" (Duijnhouwer, Prins, & Stokking, 2012, p. 171). Previous research pointed out that peer feedback enhances students' performance (eg. Falchikov, 2003). In view of formative assessment, it is rather logical that students should be given the opportunity to use this feedback, in order to improve their learning and achievement (Nicol & MacFarlane-Dick, 2004). However, many questions remain unanswered on how the formative assessment practices should be implemented into educational practice to boost students' learning in higher education (Sadler, 2010). More particular, research lacks a rigid approach on how PA practices should be tailored in function of students' learning (Strijbos & Sluijsmans, 2010). In this respect, this study is particularly focusing on how instructional interventions can customize the PA process in order to enhance students' learning.

The essence of peer feedback quality

A growing body of research emphasizes that feedback has a powerful impact on both learning and performance (Nelson & Schunn, 2008). Interestingly, the average effects of feedback are one of the highest in education, but also one of the most unpredictable in their influences (Hattie & Gan, 2011). A large body of research claims that the effectiveness of a feedback message largely depends on the content, form and function of the feedback (eg. Narciss, 2008). Especially, feedback content appears to be crucial for the impact of feedback on learning and performance (Cho & MacArthur, 2010). In literature, there is no fixed answer on what exactly determines peer feedback quality. Following the feedback framework of Hattie and Timperley (2007), high quality peer feedback should provide answers on three major feedback questions: 'Where am I going?', 'How am I going?', and 'Where to next?'. Other research emphasises on the presence of verifications and elaborations, as necessary components of high quality feedback content (Narciss, 2008). As an essential aspect of learning, verifications provide information on particular errors in performance, while elaborations attempt to help the learner in error correction. In this respect, Shute (2008) summarises in her review study on formative feedback that feedback content should be "on target (valid), objective, focused, and clear" (Shute, 2008, p. 182).

While other studies propose to examine the quality of peer feedback messages through content analysis (eg. Strijbos, Van Goozen, & Prins, 2012; Gielen & De Wever, 2015), or through the calculation of two indices, namely validity and reliability (Hafner & Hafner, 2003), prior studies have applied a scoring rubric to measure the quality of the feedback messages (Prins, Sluijsmans, & Kirschner, 2006). A scoring rubric is particularly valuable because it presents the

assessment criteria in a structured format (Panadero, Romero, & Strijbos, 2013) and it gives an indication about expected performance by listing the relevant assessment criteria and by defining the quality levels of each criterion (Andrade & Valtcheva, 2009). Prins et al. (2006) developed the Feedback Quality Index, in which a number of quality criteria are discussed. First of all, they emphasised on the importance of the assessment criteria, in which assessor and assessee are guided towards high quality performance (see also Sluijsmans, 2002). This idea is also supported by research, which claims that it is essential that the assessor is capable of identifying and understanding the assessment criteria in order to provide a reliable and valid assessment (Panadero & Jonsson, 2013). Next, students must have the knowledge and skills to sufficiently illustrate the nature of their feedback. In the FQI (Prins et al., 2006), the nature of feedback refers to specific peer feedback content such as remarks, posed questions and external examples. In this respect, feedback can be perceived as a tool to inform, motivate, or reinforce students (Nelson & Schunn, 2008), while other research made a distinction between evaluative and informative feedback in the context of PA writing assignments (Van den berg, Admiraal, & Pilot, 2006). Logically, previous research suggested that some types of feedback are more effective than others (Nelson & Schunn, 2008). In order to benefit the assessee's understanding of feedback, previous research stresses that explicit feedback features should be included or excluded (eg. Nelson & Schunn, 2008). Previous research revealed that more specific and elaborated feedback leads to improved performance and outcomes (Strijbos, Narciss, & Dünnebier, 2010). Finally, students need to be capable to transform their peer feedback in a message. According to the FQI (Prins et al., 2006), students should write their feedback in the first person throughout the whole report, in a logical and clear structure, in which short descriptions are preferable. (Prins, et al., 2006).

In an attempt to safeguard high quality peer feedback, recent research summarizes that students are involved in high-level cognitive processing during this peer feedback process (King, 2002), in which they require skills comprising "the ability to engage with and take ownership of evaluation criteria, to make informed judgements about the quality of the work of others, to formulate and articulate these judgments in written form and, fundamentally, the ability to evaluate and improve one's own work based on these processes" (Nicol, Thomson, & Breslin, 2014, p. 120). Therefore, this study is particularly interested in how we can optimize the peer feedback process with the underlying purpose to increase the feedback quality and additionally, the product score.

Structuring the peer feedback process to optimize feedback quality

As mentioned above, PA can be seen an example of a more complex learning task that requires high-level cognitive processing, however, such high-level PA processes hardly happen spontaneously (Kollar & Fischer, 2010). Previous studies pointed out the need for structure and

support to ensure effective feedback (eg. Poverjuc, Brook, & Wray, 2012). Recently, research questioned what type of support is essential for the assessor and assessee to promote high quality feedback (Hovardas, Tsivitanidou, & Zacharia, 2014). Previous research of Van Merriënboer, Kirschner, and Kester, (2003) suggested amongst others to model the use of cognitive strategies by presenting checklists and process worksheets, or by asking leading questions, in order to support students in complex learning. This type of support may be beneficial to support the roll of the assessor in providing feedback as well.

Other studies showed that structure is beneficial for the peer feedback process by, for example further specifying a peer feedback template to enhance the peer feedback quality (eg. Gielen & De Wever, 2012; Gielen & De Wever, 2015). It is within this frame that the main aim of the present study can be situated: "How can we increase the peer feedback quality by structuring the PA process?" Based on the scripted cooperation approach (O'Donnell, 1999), collaboration scripts are recommended in the literature to boost successful collaborative learning activities (Fischer, Kollar, Stegmann, & Wecker, 2013). As a script specifies, plans, and assigns roles and activities for collaborative learning activities (eg. Fischer, et al., 2013), a script can be seen as an instructional collaboration scenario (O'Donnell & Dansereau, 1992), which concentrates on socio-cognitive structuring (Kollar, Fischer, & Hesse, 2006). Since numerous contextual factors play a role, determining the accurate level of structuring appears to be the actual challenge (Dillenbourg, Järvelä, & Fischer, 2009). Therefore, this study attempts to provide an answer on how detailed the script should be and what level of structuring is the most appropriate (c.f. 'script granularity' concept of Kobbe, et al., 2007).

In order to become skilled peer assessors and assessees, who provide and receive high quality peer feedback, research stresses that students require practice and training (Sluijsmans, 2002; Birenbaum, 1996). As training is often suggested in the literature, it is important that students have the opportunity to practice similar performance at multiple occasions. For this reason and building on previous studies (eg. Gielen & De Wever, 2012), this study foresees three performance cycles. This is in line with research, which claims that students need to have the opportunity to replicate similar performance or to close the feedback loop, in order to grasp the effectiveness of the peer feedback (Boud, 2000). While performing an academic task in a wiki environment, which is praised for supporting online collaboration and assessment activities (De Wever, Van Keer, Schellens, & Valcke, 2011), this study incorporated three different feedback forms with a varying structuring degree as instructional intervention, to examine the effect on the feedback quality and product score. This instructional intervention has the purpose to activate students' internal script in an ideal situation (Fischer, et al., 2013), but previous research stressed it remains unclear which level of structure is required at a certain time in a particular context (Kobbe, et al., 2007). Taking into account under-scripting (Kirschner, Sweller, & Clark, 2006) or over-scripting effects (Dillenbourg, 2002), in which a script can be too flexible or too rigid that it eventually undermines students' learning.

Rationale for this study and expectations

It is within this frame that this study is particularly interested in to what degree the assessors' peer feedback process should be structured, in order to increase the peer feedback quality and product scores. With respect to this question, we expect peer feedback quality scores and the product scores will increase over time, as mentioned above, when learners have the opportunity to perform similar tasks at multiple measurement occasions. As students in their bachelor program habitually lack practice and experience in the peer feedback process, they may require a higher amount of structure and support, in order to become skilled peer assessors who provide high-quality peer feedback. Therefore, we expect that a higher structuring degree will lead to higher peer feedback quality scores. Consequently, we assume that students, who receive a higher degree of structuring in their peer feedback process, will have higher product scores and a higher increase from draft to final version, compared to less-structured conditions.

Methodology

Participants and procedure

The participants in the present study were first-year bachelor students Educational Sciences (N = 168), enrolled in the course Instructional Sciences that runs during the first semester of the academic year. Participants were randomly assigned to groups (n = 37) of maximum 5 students to collaborate in a wiki environment. Students had to write three times a draft and final version of an abstract of a submitted, yet not published scientific article related to the topic (ie. they received the paper, but the abstract was left out). Before writing the final version, they received peer feedback on their draft version. Each student was assigned to provide three times peer feedback (one time for every one of the three draft versions written) to one fixed specific group member with the goal to increase the quality of the final abstract. The wiki task, including the peer feedback, was part of their curriculum requirements. During the writing and assessment phase, students could access the wiki anywhere and anytime.

Research instruments

Scoring rubric for quality of peer feedback messages

First of all, the rubric to assess the peer feedback quality is based on the Feedback Quality Index (Prins et al., 2006), which is in turn based on several prior studies (eg. Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002). Following the scoring rubric that was developed to measure the quality of feedback reports of general practitioners in training (Prins et al., 2006), other previous studies (eg. Gielen & De Wever, 2012) and this particular study applied a scoring rubric, which maintains all three main categories (use of criteria, nature of the feedback, and writing style), and their involved sub categories with corresponding scoring percentages of the scoring rubric of the FQI, but focused specifically on measuring the quality of peer feedback messages of first-year higher education students. First of all, Use of criteria was categorised by the number of: used criteria, remarks per criteria, remarks focused on particular aspects of criteria, explanations of remarks per criteria, explanations of remarks focused on particular aspects of criteria.

Similar to the FQI, the use of criteria accounted for 50% of the score, in which both feedback content and explanations were assessed. Secondly, nature of feedback was categorised by the number of: positive and negative remarks, reflective questions, external examples and suggestions for improvement. These four items that identify the nature of feedback accounted in total for 35%. Finally, writing style was categorised by: structure, use of key words or descriptions, and use of first person (Prins et al., 2006). These three items defined the quality of

writing and accounted for 15%. As shown in Table 1, this resulted in a scoring rubric of 9 items with a scoring range between 0 and 100 to measure the quality of peer feedback messages.

Table 1 – Scoring rubric to measure the quality of peer feedback messages

Main category	Sub category	Good Feedback		Average Feedback		Minimal Feedback	
Criteria	Content	Comments on all feedback aspects, in combination with the associated criteria	30	Comments on some feedback aspects, in combination with the associated criteria	15	None or minimal comments	0
	Clarification	Clarification of all comments on feedback aspects	20	Clarification of some comments on feedback aspects	10	None or minimal clarification of comments on feedback aspects	0
Feedback	Comments	Equilibrium between positive and negative comments	10	Mainly positive comments	5	Mainly negative comments	0
	Asked questions	Multiple questions which stimulate reflection	10	One question which stimulates reflection	5	No asked questions present	0
	Examples	Useful examples	5	Unclear examples	2	No examples present	0
	Suggestions	Useful and concrete suggestions for future improvement; Constructive advice	10	Vague and abstract suggestions for future improvement	5	No suggestions for future improvement present	0
Writing	Structure	Clear structure	5	Unclear structure	2	No structure	0
	Formulation	Short formulations	5	Mainly keywords	2	Only keywords	0
	Style	Written in first person throughout the whole feedback message	5	Occasionally written in first person	2	No use of first person	0
TOTAL			100				

Note. Adapted from the Feedback Quality Index (Prins, Sluijsmans, & Kirscher, 2006).

Scoring rubric for quality of the wiki product

For the product score, ie. the quality of the written abstract in the wiki, a rubric was developed in which the necessary components for a good abstract are included. In academic writing, literature refers to components such as intention of the study, problem statement, methodology, findings and conclusions, limitations, structure, language, etc. Therefore, this study developed a scoring rubric in which these components are categorised in four main categories (situating the study, content of the abstract, style, and general impression) and nine

corresponding sub categories. First of all, situating the study was categorised by how well described are: the intention or focus of the study, the context of the problem statement, and finally the continuity between the focus of the study and the context of the problem statement. These three items accounted for 30% of the total score. Secondly, the content of the abstract was categorised by the methodology with corresponding details on the setting, the results being all present and concisely formulated, and finally the presence of limitations and suggestions for future research. These three items accounted for 25% of the total score. Thirdly, the main category style was categorised by: structure of the abstract, language and writing style, and finally word count. These three items accounted for 25% of the total score. Finally, the main category 'General impression' was categorised by the impression of completed effort and corresponding need for revision. This main category accounted for 20% of the total score. Therefore, the developed scoring rubric to analyse the quality of the wiki product had a scoring range between 0 and 100.

Conditions

The instructor provided a peer feedback form for each of the three conditions, presented as a template with a list of ten criteria (intention of research, problem statement, methodology, results, conclusion, limitations, structure, language, deadline, and general judgment). This list of criteria was submitted to the no structure condition, but students in this condition received no further instructions, while the two other conditions received additional instructions. The basic structure condition received additionally two guiding questions ('What was good about your peers' work?' and 'What would you change in your peers' work?'). Students in the elaborate structure condition received a template, which was structured according the principles of feed up, feedback, and feed forward (Hattie & Timperley, 2007) and additionally in each of these three sections, the list of criteria was simply repeated.

Data analysis

Given the clear hierarchical structure of the data, namely three measurement occasions (i.e. the peer feedback moments, indicated by the variable 'time', level 1) are nested within each of the 168 students (level 2), who are in turn nested within 37 groups (level 3), multilevel modelling (MLwiN 2.29) was used to analyse the peer feedback quality and the product quality (ie. the quality of the versions of the abstract written in the wiki). Initially, for both peer feedback score and product score a fully unconditional null model was tested to examine whether a multilevel approach was required compared to a single-level regression analysis. Next, the categorical predictor time was added to the null model, which resulted in a compound symmetry model, which is a random intercept model with no explanatory variables except for

the measurement occasions (Snijders & Bosker, 1999). In this model, the two last measurement occasions (time 2 and 3) can be compared with the reference category (time 1). After this, the followed procedure is dissimilar for the peer feedback and product score. Regarding the peer feedback score, the categorical predictor 'condition' is added in the next step. In a final phase, the interaction condition*time was added to the model. Regarding product score, first the categorical predictor 'version' was added to the model, as the product score has two versions, namely draft and final version. After this, the categorical predictor condition is added in a next phase. Finally, the interaction time*condition was added to the model. Further exploration of other interaction possibilities revealed no significant interaction effects when version was involved and was therefore excluded from the multilevel model. By using a stepwise multilevel approach, the additional value of each subset of variables to the model could be checked.

Hypotheses

With respect to the quality of the peer feedback and product, the following hypotheses are put forward:

(H1) Over time, students in all three conditions will have significantly higher peer feedback quality scores, more specifically (H1a) from time 1 to time 2, (H1b) from time 2 to time 3, and (H1c) from time 1 to time 3.

(H2) Students will have higher peer feedback quality scores, more specifically (H2a) the basic structure compared to the no structure condition, (H2b) the elaborate structure compared to the no structure condition, and (H2c) the elaborate structure compared to the basic structure condition.

(H3) Over time, students in all three conditions will have significantly higher product quality scores, more specifically (H3a) from time 1 to time 2, (H3b) from time 2 to time 3, and (H3c) from time 1 to time 3.

(H4) Students will have higher product quality scores, more specifically (H4a) the basic structure compared to the no structure condition, (H4b) the elaborate structure compared to the no structure condition, and (H4c) the elaborate structure compared to the basic structure condition.

(H5) The product quality scores improve significantly better from draft to final version for students, more specifically (H5a) for all conditions, no matter what structuring degree they receive (main effect), but even more in (H5b) the basic structure compared to the no structure condition (interaction effect), (H5c) the elaborate structure compared to the no structure condition (interaction effect), and (H5d) the elaborate structure compared to the basic structure condition (interaction effect)

Results

Peer feedback score

All models were created following the previously described stepwise procedure and are represented in Table 2. The random-intercept three-level null model (Model 0) predict the overall peer feedback score across all feedback moments (time), students, and groups (the intercept; ie. 53.23 out of 100). The null model divides the variance of peer feedback scores into between groups, within groups between students, and within students between peer feedback moments. The results show that 19.32% of the total peer feedback variance is situated at the group level (p=.002), the proportion of variance due to differences between students within groups was 11.23% (p=.014), and finally 69.45% of the total variance is situated at the time level (p<.001, see Table 2).

Next, the categorical predictor time was added to the null model, which resulted in Model 1. Adding this variable to the null model resulted in a better model fit ($\chi2$=98.309, df=2, p<.001). The results presented in Model 1 reveal a significant effect of measurement occasion on peer feedback scores, indicating that the quality of the feedback was significantly higher the second and the third moment (compared to the first moment, both at p<.001). On average, the quality of the feedback was also significantly higher the third moment compared to the second (p=.032). Following, the categorical predictor condition was added to Model 1, which resulted in Model 2. The condition in which students did not receive any additional structure in their peer feedback template was taken as reference category. Adding this variable resulted in a better model fit ($\chi2$=13.308, df=2, p=.001). In the last step, the interaction effects between time and condition were added. However, as this model did not result in a better fit than Model 2 ($\chi2$=1.605 df=4, p=.808) and none of the interaction effects were significant, Model 2 was chosen as final model for further analysis.

Table 2 – Multilevel models for the quality of the feedback (dependent variable: peer feedback score)

	Model 0	Model 1	Model 2 (final model)
Fixed			
Intercept (cons)	53.231(1.783)***	43.317(2.019)***	40.163(2.747)***
Time 2		13.119(1.640)***	13.119(1.640)***
Time 3		16.625(1.640)***	16.625(1.640)***
Basic structure			-1.717(3.601)
Elaborate structure			11.790(3.675)**
Random part			
Level 3 - Group	84.202(27.625)**	84.202(27.625)**	48.038(19.326)*
ρ(%)	19.32%	21.89%	13.77%
Level 2 - Student	48.961(20.086)*	74.561(19.406)***	74.864(19.442)***
ρ(%)	11.23%	19.38%	21.46%
Level 1 - Time	302.750(23.358)***	225.951(17.433)***	225.951(17.433)***
ρ(%)	69.45%	58.73%	64.77%
Model fit			
Deviance (-2LL)	4422.783	4324.474	4311.166
χ^2		98.309	13.308
df		2	2
p		p<.001	p=.001
Reference model		Model 0	Model 1

Notes: *p < .05. **p < .01. ***p < .001.

In Model 2, the results indicate a significant main effect of measurement occasion and condition, with respect to the peer feedback scores, as shown in Figure 1. First of all, peer feedback scores increased significantly over time for all students in all groups, both significantly from time 1 to time 2 with an increase of 13.12 (p<.001), and from time 2 to time 3 with an increase of 3.51 (p=.033), causing a total increase from time 1 to time 3 of 16.625 (p<.001). These findings confirm respectively H1a, H1b and H1c. Secondly, results point out that students who received an elaborate structure have an overall significantly higher peer feedback score, which is in more detail 11.79 higher compared to the no structure (p=.001), and 13.51 higher than the basic structure (p<.001) condition, confirming H2b and H2c. Between the no structure and basic structure no significant differences were found (p=.190), not supporting H2a.

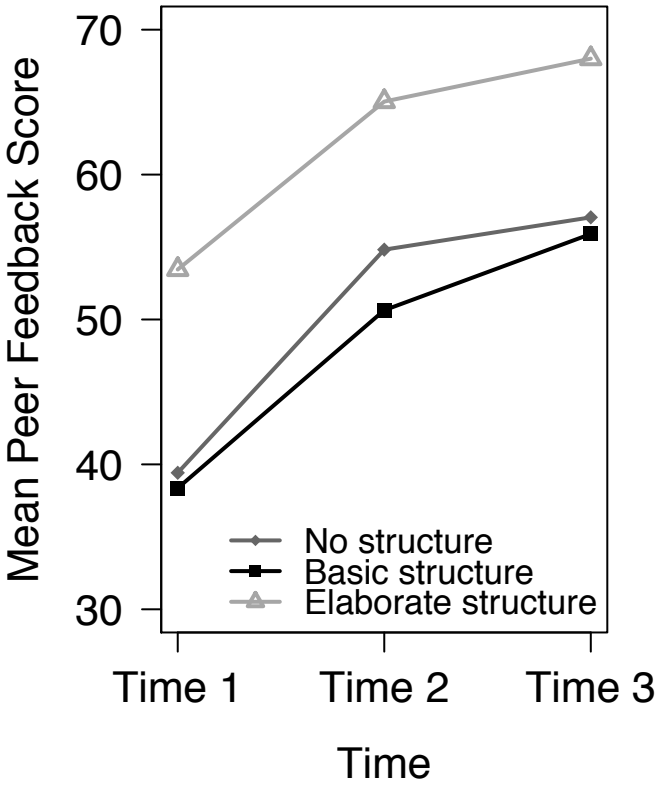


Figure 1. Peer feedback scores over time

Product score

The null model showed that 1.02% of the total peer feedback variance is situated at the group level (p=.478), the proportion of variance due to differences between students within groups was 5.58% (p=.040), and finally 93.40% of the total variance is situated at the time level (p<.001, see Table 2). After estimating the null model, the categorical variables time and version were added to the null model as measurement occasions (Model 1). The results presented in Model 1 reveal a significant main effect of measurement occasion on product scores over time. Adding these two variables resulted in a better model fit (χ2=838.695, df=3, p<.001). After estimating Model 1, interaction effects between time and versions were checked for, but no significant effects were found, indicating that the increase in score between the draft version and the final version was about the same at each of the three moments. For Model 2, the categorical predictor condition was added, and the results revealed a significant main effect. Model 2 did not fit the data better than Model 1 (χ2=0.626, df=2, p=.731) and no main effect of condition was found (see Model 2, Table 2). After estimating Model 2, interaction effects between condition and version were checked for but not found. However, in a next step, the interaction effects of time and condition were added to Model 2, resulting in Model 3, and revealing significant interaction effects. Model 3 also fitted the data better than both Model 2 (χ2=20.884 df=4,

p=.001) and Model 1 (χ2=21.510, df=6, p=.001), and was therefore chosen as final model for further analysis.

In Model 3, results indicate a significant main effect of the two categorical predictors time and version, with respect to the peer feedback scores. Firstly, results show that the product scores increased significantly over time for all students, confirming H3a, H3b and H3c. More specifically, the product scores improved significantly from time 1 to time 2 with an increase of 22.439 (p<.001), as well as from time 2 to time 3 with an increase of 7.18 (p<.001), causing a total increase from time 1 to time 3 of 29.61 (p<.001). Secondly, results point out that the product score increased significantly from draft to final version for all students, confirming H5a, with an average increase of 9.14 (p<.001). No interaction effects were found between condition and version and therefore H5b, H5c and H5d are not supported.

Table 3 - Multilevel models for the quality of the wiki product (dependent variable: product score)

	Model 0	Model 1	Model 2	Model 3 (final model)
Fixed				
Intercept (cons)	63.864 (.862)***	39.329(1.111)***	38.452(1.637)***	41.059(1.814)***
Time 2		24.929 (1.066)***	24.929(0.993)***	22.439(1.684)***
Time 3		34.946 (1.066)***	34.946(0.993)***	29.614(1.684)***
Final version		9.139(0.811)***	9.139(0.811)***	9.139(0.801)***
Basic structure			1.637(2.077)	-0.699(2.488)
Elaborate structure			0.980(2.123)	-4.728(2.544)
Time 2 . Basic				1.165(2.372)
Time 3 . Basic				5.843(2.372)***
Time 2 . Elaborate				6.618(2.426)**
Time 3 . Elaborate				10.509(2.426)***
Random part				
Level 3 - Group	4.894(6.901)	4.894(6.901)	4.823(6.862)	4.823(6.862)
ρ(%)	1.02%	1.60%	1.97%	2.00%
Level 2 - Student	26.861(13.088)*	74.864(19.442)**	73.881(12.601)***	74.560(12.597)***
ρ(%)	5.58%	* 24.49%	30.22%	30.93%
Level 1 - Time	449.892(21.952)**	225.951(17.433)*	165.763(8.088)***	161.693(7.890)***
ρ(%)	* 93.40%	** 73.91%	67.81%	67.07%
Model fit				
Deviance (-2LL)	9077.192	8238.497	8237.871	8216.987
χ2		838.695	0.626	20.884
df		3	2	4
p		***	p=.731	***
Reference model		Model 0	Model 1	Model 2

Notes: *p < .05. **p < .01. ***p < .001.

Regarding condition, multilevel analysis revealed no main effect, contradicting H4a, H4b and H4c. Though when taking into account the feedback moments, results showed an interaction-effect, suggesting that the product scores evolve significantly different over time for particular

conditions, as shown in Figure 2. More specifically for time 1, results show that the product score of the basic condition was 0.70 lower compared to the no structure (p=.778). Students in the elaborate structure condition had a lower product score at the start, which is in more detail 4.73 lower compared to the no structure (p=.063), and 4.03 lower compared to the basic structure condition (p=.114) at time 1.

Regarding time 2, results reveal that students who received an elaborate structure, have a slightly higher (but not significant) product score, which is in more detail 1.89 higher compared to the no structure (p=.457), and 1.42 higher compared to the basic structure (p=.573) condition. Also the product score of the basic structure was 0.47 higher compared to the no structure condition (p=.851) at time 2. Finally for time 3, results show that students who received no structure have an overall significant lower product score, which is in more detail 5.14 lower compared to the basic structure condition (p=.039), and 5.78 lower than the elaborate structure condition (p=.023). This only partly (i.e. only at time 3) confirms H4a and H4b. The product score of the basic structure was 0.64 lower compared to the elaborate structure condition (p=.802), so H4c is not supported.

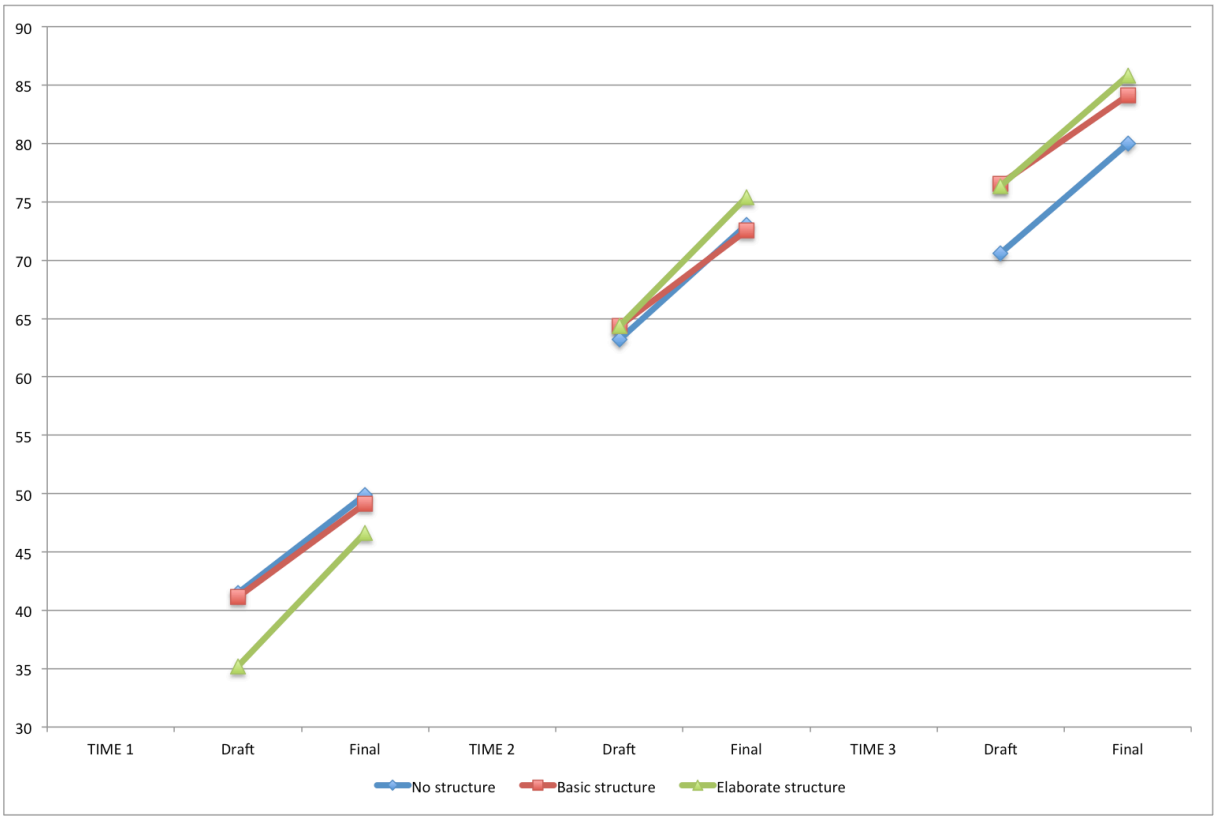


Figure 2. Wiki product scores over time

In sum, Figure 2 represents the main findings clearly. Firstly, the progress from draft to final version is almost equal for all conditions at all moments (see the similar slopes in Figure 2). Over

time, product scores improve overall, but point out no differences between the conditions. However, if we look closer at each moment, results show that at time 1 the elaborate structure has lower (but only nearly significant) product scores than the two other conditions, while at time 2 the elaborate structure already has slightly higher, but not significant higher product scores compared to the less structured conditions. Interestingly, at time 3, both the elaborate and basic structure condition have significantly higher product scores compared to students who receive no additional structure in their feedback process.

Discussion

This study examined how the degree of structuring the peer assessment process has an impact on the feedback and product quality, when students compose feedback with the help of a peer feedback form with a varying structuring degree. Finally, the practical implications and direction for future research are presented.

Peer feedback quality

Over time, the results revealed that the feedback scores increase significantly for all conditions, suggesting that students overall provide peer feedback messages of a better quality, when they gain experience through practice. Although a few studies claim that students can offer valuable feedback without actual training in assessment (eg. Cho & MacArthur, 2010), other research advocates that students benefit from practice and training in receiving and providing peer feedback (eg. Sluijsmans, 2002) and moreover, that students require practice to become skilled peer assessors (eg. Birenbaum, 1996; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). Measured by the Feedback Quality Index (Prins et al., 2006), the results of the present study revealed that students in the elaborate structure condition have significantly higher feedback quality scores, compared to students who received merely some guiding questions or who received no additional structure at all. This is in line with previous experimental studies, which claim that structure is beneficial for the peer assessment process (eg. Gielen & De Wever, 2012) and, which underline the need for structure and support to ensure effective feedback (eg. Poverjuc, Brook, & Wray, 2012). More specifically, offering students a peer feedback form including a criteria-oriented list structured according the three feedback principles feed up, feedback and feed forward (Hattie & Timperley, 2007) appears to be an effective approach to increase significantly the peer feedback quality. As finding the accurate level of scripting is the actual challenge (Kobbe, et al., 2007), we believe that further scripting the peer feedback process by providing an elaborate structure in a peer feedback template is favourable approach, to enhance the peer feedback quality (eg. Fischer, Kollar, Stegmann, & Wecker, 2013).

Product quality

The results indicated an overall significant increase over time for all students, no matter what level of structure they receive in their peer feedback process. Previous research claims students' learning can be facilitated by engaging students actively in PA (eg. Li, Liu, & Steckelberg, 2010), as PA has several cognitive gains for both assessor and assessee, such as increased attention on the crucial elements, which determine high quality work (Topping, 1998). Following, results demonstrated an overall significant increase of product scores of 9% from draft to final version. This is in line with research, which underlines that feedback can have a large impact on performance (Nelson & Schunn, 2008), as it "might also reveal the next small steps needed to improve quality" (Topping, 1998, p. 255). This is supported by a review, which advocates that every variety of feedback, whatever its amount or specificity, can have a positive effect on students' product scores (Topping, 1998).

With respect to the provided level of structure in the peer feedback process, overall results revealed no significant differences between the conditions regarding product quality scores. However, when taking a closer look, interaction effects pointed out some significant differences between the conditions over time. In general, research advocates that the quality of students' performance increases over time, whenever they have the opportunity to practice similar learning activities (eg. Sluijsmans, 2002). More specifically in this study at time 3, results indicated that students of both basic and elaborate structure conditions had higher product quality scores after multiple practice occasions, compared to students who did not receive any additional structure in the peer feedback process. These findings suggest that structure in the peer feedback process has the potential to boost product scores, while it is important that students use this feedback, in order to improve their performance (Nicol, & MacFarlane-Dick, 2004). This is supported by other research, which advocates that structure, in which the roles and activities of involved learners are further concretised, can be valuable for students' learning (Schellens & Valcke, 2006).

To conclude, our results indicate that further scripting the peer feedback process is beneficial for the quality of students' peer feedback and product performance, which is in line with similar previous studies (eg. Gielen & De Wever, 2012). Over time, all students improved significantly after multiple practice occasions in providing peer feedback and finishing their wiki task. It became clear that in the end students, who received an elaborate degree of structure to provide peer feedback, had significantly higher peer feedback quality scores compared to less structured conditions. Furthermore, students who received additional structure in their peer feedback template in the end had significantly higher product quality scores after similar practice occasions, compared to students who did not receive any additional structure in the peer feedback process. Therefore, this study advocates that offering additional structure in PA, to further specify the role of the assessor during the peer feedback process, is a valuable approach to increase both the quality of peer feedback and performance.

Limitations, directions for future research and practical implications

Due to various contextual factors, it is not possible to determine if the increase in product scores from draft to final version is caused by actually receiving or providing peer feedback, or merely because students had another opportunity to revise their wiki product once more, with or without taking into account the peer feedback. Future studies could close the feedback loop (Boud, 2000), in which the assessee could be structured to evaluate the received peer feedback after revision. Moreover, the present study incorporated the FQI to measure the quality of peer feedback messages, while in-depth content analysis could be another approach to determine the actual peer feedback quality, as it provides a more detailed insight of the specific peer feedback content. Another direction for future research could be to examine the added value of a structured peer feedback form in various educational contexts and over a longer period of time. A last suggestion for future research could be to examine how the role of the assessee could be structured as well, by for example a peer feedback request, whilst the majority of the experimental studies in the literature focus merely on the assessor (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

Instructors wishing to implement peer assessment should consider the following two recommendations. Firstly, this study recommends implementing a peer feedback template with a higher structuring degree, as instructional intervention to support students during the peer feedback process. Therefore, this study proposes that a peer feedback template should consist out of two essential features. On one hand, the template needs to provide a list of the pre-specified, or preferably mutual discussed criteria (Sluijsmans, 2002). On the other hand, this template could be inspired by feedback framework of Hattie and Timperley (2007), in which students are encouraged to provide feedback on past performance and feed forward in function of future performance, focused on particular criteria. Secondly, this study supports a large body of research that encourages instructors to foresee multiple practice occasions, in which students are involved in peer assessment and similar task performance.

References

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice, 48*, 12–19.

Birenbaum, M. (1996). Assessment 2000: towards a pluralistic approach to assessment. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). Boston, MA: Kluwer.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 7-74.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151–167.

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328–338.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14*, 201-206.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen: Open University of the Netherlands.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research in computer-supported collaborative learning: from design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3-19). Springer.

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction, 22*, 171–184.

Dysthe, O. (2004). *The challenges of assessment in a new learning culture*. The 32nd International NERA/NFPF Conference, Reykjavik, Iceland.

Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 157-166). London: Kogan Page.

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching, 3*, 102–108.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational psychologist, 48*, 56–66.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585–594.

Gielen, M., & De Wever, B. (2015). Structuring Peer Assessment: Comparing the impact of the degree of structure on peer feedback content quality.

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.

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education, 25*, 1509–1528.

Hattie, J. & Gan, M. (2011). Instruction based on feedback. In Mayer, R. E. & Alexander, P. (eds). *Handbook of research on learning and instruction* (pp. 249-271). NewYork: Routledge. Taylor and Francis Group.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of educational research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152.

King, A. (2002). Structuring Peer Interaction to Promote High-Level Cognitive Processing. *Theory Into Practice, 41*, 33-39.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., Hakkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 211-224.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - A conceptual analysis. *Educational Psychology Review, 18*, 159–185.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344–348.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*, 525-536.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125-143). Mahwah, NJ: Erlbaum.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401.

O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In: A. M. O'Donnell, & A. King (eds.), *Cognitive perspectives on peer learning*. (pp. 179–196). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Donnell, A. M., & Dansereau, D. F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 120–141). Cambridge, MA: Cambridge University Press.

Nicol, D.J. and Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice. In C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. Ross & B. Smith (Eds.) *Enhancing Student Learning Through Effective Formative Feedback* (pp. 3-14). York: The Higher Education Academy.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*, 102-122.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129-144.

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*, 195-203.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465–477.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–44.

Sadler, D.R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*, 535–550.

Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education, 46*, 349–370.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153 – 189.

Sluijsmans, D. (2002). *Student involvement in assessment: the training of peer assessment skills*, unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.

Sluijsmans, D., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education, 27*, 443–454.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

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.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages.* Paper presented at the meeting of EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

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

Topping, K .J. (2009). Peer Assessment. *Theory Into Practice, 48*, 20–27.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. *Assessment and Evaluation in Higher Education, 31*, 9-16.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Berg, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*, 316-327.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind : Instructional Design for Complex Learning. *Educational Psychologist, 38*, 5–13.

4 Structuring Peer Assessment: Comparing the Impact of the Degree of Structure on the Peer Feedback Content

This chapter is based on:

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315-325.

Chapter 4
Structuring Peer Assessment: Comparing the Impact of the Degree of Structure on the Peer Feedback Content

Abstract

The present study examines the added value of structuring the peer assessment process, by providing students with a peer feedback template with a varying structuring degree, for the peer feedback content quality in a wiki environment in higher education. The present study took place in the 1st year of a university course in Instructional Sciences (N=176) and more specifically compared three conditions: no structure peer feedback (control), basic structure peer feedback, and elaborate structure peer feedback condition. Quantitative content analysis of students' (n=41) peer feedback messages was performed, and analyses of (co)variance revealed some discrepancies between the conditions regarding the proportion of peer feedback content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus. This study demonstrated that a higher structuring degree in a peer feedback template during the peer assessment process can have an impact on peer feedback content with respect to the abovementioned categories the peer feedback content. Results revealed significant differences between the three conditions regarding the peer feedback content categories.

Introduction

A large body of research underlines the power of assessment *for* the learning process (Evans, 2013; García, García-Álvarez, Moreno, 2014; Kennedy, Chan, Fok, & Yu, 2008; Pellegrino, Chudowsky, & Glaser, 2001). The shift from 'assessment of learning' towards 'assessment for learning' requires learners to be actively involved in all phases of the assessment process (Dysthe, 2004; Boud & Molloy, 2013). Assessment gives learners an indication of not only their strengths and weaknesses, but also of the next steps to be taken in the learning process. Therefore, the value of implementing more formative assessment approaches in education – in order to answer the call for more assessment *for* learning – have been advocated widely in the literature (e.g. Black & William, 1998, Sadler, 1989, Strijbos & Sluijsmans, 2010). However, many questions remain unanswered on how the formative assessment practices should be implemented into educational practice to boost students' learning in higher education (Sadler, 2010). As a common method of formative assessment, peer assessment (PA) has demonstrated its educational value for learning (see e.g. Topping, 2010). More particular, the educational potential of online PA for students' learning has been widely discussed (eg. Cheng, Liang, & Tsai,

2015). In this respect, research has shown that involving learners in online PA activities appeared for example to have an advantageous effect on students' writing performance (eg. Gielen & De Wever, 2015). Yet, research on PA in higher education is up to now "very variable in type and quality, scattered and fragmentary in nature" (Topping, 1998, p. 267; see also Evans, 2013, who still presents the same conclusion). When being involved in formative PA practices, the assessor needs to be proficient to deal with specific assessment criteria, evaluate a peer's performance and finally, compose a valuable peer feedback message. On the other hand, the assessee needs to be capable to question the assessor's peer feedback and makes changes accordingly, where the assessee is willing to follow the assessors' advice, in order to augment the quality of the performance (Hovardas, Tsivitanidou, & Zacharia, 2014). Previous research stresses that PA practices require more 'constructive alignment' (Biggs, 1996), in which specific PA practices should be intentionally tailored in function of expected students' learning (see also Strijbos & Sluijsmans, 2010).

When we examine earlier research on PA, we can notice that within the field of PA, especially peer feedback is often seen as an important educational practice of PA (e.g. Falchikov, 1995). Also, other review studies identify peer feedback as a constructive technique for enhancing students learning (e.g. Topping, 1998), such as enhancing the quality of the students' writing (Van Zundert, Sluijsmans, & Van Merriënboer, 2010). Previous research illustrated that peer feedback on the social performance of individual group members can increase the performance and attitudes of a CSCL-group (Phielix, Prins, & Kirschner, 2011). However, research on the impact of peer feedback on students' learning is lacking (Hattie & Timperley, 2007). Although there is some research that indicates that feedback content appears to play an essential role (e.g. Cho & MacArthur, 2010), detailed studies on how different peer feedback content is influencing students' activities is lacking (Strijbos, Narciss, & Dunnebier, 2010). For this reason, research has advocated that all responsible actors such as instructors and researchers should attempt to shed more light on the required type of structure and support an assessor needs for compiling high quality peer feedback (Hovardas, et al., 2014).

Therefore, the present study wants to examine the content of peer feedback in detail. More specifically, this study builds further on an earlier study (Gielen & De Wever, 2015) in which the added value of different peer feedback forms, with a varying degree of structuring, was studied in a wiki environment in higher education, with respect to product scores. Also, a general peer feedback quality index (Prins, Sluijsmans, & Kirschner, 2006) was used to assess the content quality of peer feedback messages. However, the content of the peer feedback was not analysed in detail. In order to study the peer feedback content quality in more detail, the present study was set up, in which a developed content analysis scheme (which will be further discussed in this article) for exploring the specific peer feedback content quality.

Peer assessment for learning: Peer feedback as an educational practice

With regard to assessment *for* learning, formative assessment is "specifically intended to provide feedback on performance to improve and accelerate learning" (Sadler, 1998, p. 77). Feedback can be perceived as a practice of formative assessment, which attempts to close the gap between current and desired performance (Sadler, 1989). As an embraced method of formative assessment, PA has been attributed a lot of potential (Black & William, 1998). In this respect, a continuously growing body of research pointed out the value of PA both as an assessment tool (e.g. Cheng & Warren, 1997) and as a learning tool (e.g. Topping, 1998). PA challenges learners in providing feedback on a peer's performance and in receiving feedback from a peer on one's own performance. However, we cannot assume that all students will be competent to offer high quality feedback for several reasons such as proficiency (eg. Cheng, Liang, and Tsai, 2015). In this respect, previous research emphasised on the fact that students will require unique skills to proficiently perform their role as assessor and assessee (Hovardas, et al., 2014). More specifically, learners develop skills to compile judgments about the quality of a peer's work, based on specific expectations of high-quality work (Topping, 1998). Based on this, the present study focuses on peer feedback as an educational approach of PA.

Following Hattie and Timperley (2007), in order to enhance learning when there is a discrepancy between what is understood and what is aimed to be understood, feedback should provide answers on three major feedback questions: 'Where am I going?', 'How am I going?', and 'Where to next?'. To improve performance, previous research has emphasised on identifying which feedback features should be included or excluded to benefit the understanding of feedback (e.g. Nelson & Schunn, 2008). Feedback content appears to be crucial for the impact of peer feedback on learning and performance (e.g. Cho & MacArthur, 2010). Related to this, earlier research investigated simple versus elaborated feedback (Narciss, 2006; 2008) and concise general versus elaborated specific feedback (Strijbos, et al., 2010). Topping (2010) comments that elaborated and specific feedback leads to better performance. Although a growing body of research claims that feedback has a powerful impact on both learning and performance (e.g. Nelson & Schunn, 2008), a review study revealed recently that more research on the *impact of peer feedback on learning and performance is needed* (eg. Evans, 2013).

Peer feedback content

Previous literature highlights that the quality of a feedback message is determined by its content, template, and function (Narciss, 2006, 2008; Narciss & Huth, 2004; Shute, 2008). As the power of peer feedback heavily depends on its content (e.g. Cho & MacArthur, 2010), it is important to reflect on what exactly defines peer feedback content quality. In earlier studies, the developed Feedback Quality Index (Prins, Sluijsmans, & Kirschner, 2006) was incorporated to measure the quality of feedback, with the help of a scoring rubric (e.g. Gielen & De Wever, 2012;

Gielen & De Wever (2015). In the present study, however, the aim was to take a closer look at the peer feedback content and more specifically at the peer feedback style, type, and focus of messages that peers provide to each other during writing assignments in a wiki-based CSCL environment. Following, these categories will be discussed in further detail.

With regard to the *peer feedback style*, a growing body of research suggests that the content of an effective feedback message should provide two types of information: verification and elaboration (Kulhavy and Stock, 1989; Narciss, 2008), and preferably includes both elements (e.g. Bangert-Drowns, Kulik, Kulik, Morgan, 1991; Mason & Bruning, 2001). In this study, we will distinguish between verification and elaboration and a third category "general", which refers to general statements that can be labelled as neither verification nor elaboration. Verification can be described as "a dichotomous judgment to indicate that a response is right or wrong" (Hattie & Gan, 2011, p. 253) and an elaboration is the component of the feedback message, which "contains relevant information to help the learner in error correction" (Hattie & Gan, 2011, p. 253). Complementary to peer feedback style, we discuss the category *peer feedback type* for both verifications and elaborations, as students require feedback that tells them not only if they dealt with particular criteria correctly or not, but also why and what they should do about it to improve (eg. Coll, Rochera, & De Gispert, 2014). Related to this, previous research revealed that offering additional informational feedback, which justifies a particular evaluation, is beneficial for students' performance (Walker, 2015). For this reason, a balanced proportion of verifications and elaborations may be more beneficial. Following Strijbos, Van Goozen, and Prins (2012), we will distinguish between positive, negative, and neutral verifications. This is in agreement with research that claims that feedback can be positive, negative, or neutral (Topping, Smith, Swanson, & Elliot, 2000). Although assessees are more happy with positive than with negative feedback (eg. Anseel & Lievens, 2006), previous research points out that both positive and negative feedback can have a major influence on learners' performance (Podsakoff & Farh, 1989), as it can lead to a rise or drop in effort and goal setting (e.g. Bandura & Cervone, 1986). Related to this, research revealed that praise improves motivation with low-performers, but not with high-performers (Mumm & Mutlu, 2011). When learners receive negative feedback, this could lead to "giving up", but as well to "trying harder". Similarly, when learners receive positive feedback, this could result in "sitting on their laurels", but as well in "doubling their efforts" (Van Dijk and Kluger, 2004). This is in line with earlier research, which claims that both positive and negative feedback can have positive outcomes for students' learning (Kluger & Denisi, 1996).

With respect to the types of elaboration, research claims "feedback elaboration has even more variations than verification" (Shute, 2008, p.158). Rogers (1951) claimed that feedback could be evaluative, interpretative, supportive, probing, and understanding, while in another study four different feedback attitudes are mentioned, namely, authorative, interpretive, probing, and collaborative (Lockhart & Ng, 1995). More recent research perceived feedback as informational, motivational, or reinforcing (Nelson & Schunn, 2008). Van den berg, Admiraal, and Pilot (2006) made in this respect a distinction between evaluative and informative feedback in the context of PA writing assignments. Inspired by this research, the present study proposes to divide between informative and suggestive elaborations, comparable to the concepts of

feedback and *feed forward* (Hattie & Timperley, 2007). Therefore, we will differentiate between informative and suggestive elaboration. Informative peer feedback gives more details about previous performance without giving feed forward, while suggestive peer feedback specifically elaborates on how future performance can be improved. With regard to the *focus of peer feedback*, this can be specific and directive, such as addressing an error, topic or response, or on the other hand be general and facilitative, such as providing guidance or worked examples (Shute, 2008). Directive feedback aims to inform the learner about what needs to be revised exactly and is more specific than facilitative feedback in which comments and suggestions are made to support peers in their revision (Black & William, 1998). Regarding the verification and elaboration focus, the present study examines if the peer feedback is general or specific, and if the focus addresses the overall performance, particular criteria or language aspects.

To sum up, this study takes a closer look at the specific content of peer feedback messages, which students provide to each other during wiki tasks, by categorising the peer feedback content according to five main categories namely peer (1) feedback style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus. As feedback content appeared to be essential for its effectiveness, an intervention was set up to enhance the content of peer feedback messages by structuring the PA process, more specifically by further specifying the role of the assessor. In order to explore what type of support is required for the assessor to promote high quality feedback (Hovardas, et al., 2014), the present study aims to investigate the effect of this intervention, through analysing the content of the feedback.

Scripting PA to augment peer feedback content quality

PA can be seen an example of a more complex learning task that requires high-level cognitive processing, however, such high-level PA processes hardly happen spontaneously (Kollar & Fischer, 2010). Literature recommends the use of collaboration scripts to enhance successful collaborative learning activities (Fischer, Kollar, Stegmann, & Wecker, 2013). While other research emphasized on the importance of effective group formation in a collaborative environment (eg. Vargas-Vera, Nagy, & de Pablos, 2013), collaborative learning can be seen as an instructional strategy whereby students at different performance levels work together in small groups to accomplish a common learning goal (Dillenbourg, 1999). The main aim of *scripting* is to "trigger engagement in social and cognitive activities that would otherwise occur rarely or not at all" (Kobbe, et al., 2007, p.212). Scripts are not merely focused on gaining domain-specific knowledge, but also on obtaining the necessary skills to perform the scripted collaborative activities (Wecker & Fischer, 2007). Grounded in the scripted cooperation approach (O'Donnel, 1999), a script can be perceived as an instructional collaboration scenario (O'Donnell & Dansereau, 1992), which focuses on socio-cognitive structuring (Kollar, Fischer, & Hesse, 2006) by specifying, scheduling, and delegating roles and activities for collaborative learning activities (e.g. Fischer, et al., 2013). Previous research claims that role assignment is an essential structuring tool to increase knowledge construction in asynchronous discussion groups (De

Wever, Van Keer, Schellens, & Valcke, 2010). It is within this frame that the main aim of the present study can be situated (see also Gielen & De Wever, 2015): "How can we increase the peer feedback quality by structuring the PA process?"

With respect to this question, suggestions have been made in the literature. As previous research has illustrated how structuring can be an effective strategy to improve both students' PFB quality and performance in function of enhancing the actual PA process (Gielen & De Wever, 2015), the instructor could structure the PA process by providing more detailed instructions on expected performance (Kollar, Fischer, & Slotta, 2007), e.g. by providing guiding questions to support the assessor while providing peer feedback (Gielen & De Wever, 2012). One remaining question, however, is how detailed the script should be and what level of structuring is the most appropriate (c.f. 'script granularity' concept of Kobbe, et al., 2007). Although scripting can be seen as an ideal way to stimulate collaborative processes, earlier research also warned us for an "over-scripting" effect (Dillenbourg, 2002), in which a script can be so rigid that it results in less – instead of more – efficient collaboration (Fischer, et al., 2013). As research on both high and low structured scripts is growing, literature reveals that determining the accurate level of structuring is the actual challenge (Dillenbourg, Järvelä, & Fischer, 2009), as various contextual factors play a role. Recent research claims that finding the right level of scripting depend relatively on the structure of the learners' own internal script (Fischer et al., 2013). For this reason, research has advocated that all responsible actors such as instructors and researchers should attempt to shed more light on the required type of structure and support an assessor needs for compiling high quality peer feedback (Hovardas, et al., 2014). In the context of the present study, we especially want to find out more about to *what degree students' PA process should be structured to have an impact on the peer feedback content*.

Research aim

The main aim of the present study is to research how peer feedback, as an educational approach of PA, can have an impact on students' learning and performance, and how we can increase this potential impact by scripting the process. In more detail, we will study the peer feedback content quality and investigate which amount of structuring is most appropriate. For this reason, this study contributes to the current research as it examines whether shifting the structuring degree of the peer assessment process has an impact on students' peer feedback content. As explained in the theoretical framework, the effectiveness of feedback heavily depends on the actual content and thus it is important to examine practical instructional interventions in the feedback process, which have the potential to increase the quality of the feedback content that assessors offer in their PFB messages. Like this, this research can shed more light on the actual effect of providing the assessor a varying structuring degree during the PA process on the PFB content quality in a wiki environment in higher education. More specific hypotheses are formulated later in the methodology section.

Material and methods

Participants and procedure

The participants in the present study were first-year bachelor students in Educational Sciences (N = 168), who were enrolled in the course Instructional Sciences, which is a major introduction course, accounting for 7 ECTS, obliged for all students in Pedagogical Sciences. This course runs during the first semester of the academic year and more particular in the academic year 2012 – 2013 at the university of Ghent. Participation was part of their curriculum requirements, in which participants were randomly assigned to groups (n = 37) of maximum 5 students to collaborate on one wiki. During the writing and assessment phase, students could access the wiki anywhere and anytime. a bullet-pointed criteria list.

The complete wiki assignment lasted for nine weeks, in which three cycles of three weeks each were organized. Within each cycle, students were asked to write an abstract of a scientific article in three phases. Each individual student within a group was provided with a submitted, but not yet published scientific article, for which the abstract was removed, meaning that they had access to the full body of the scientific article, but had to write the abstract themselves. This was done in three phases: (1) writing a draft version of the abstract (2) providing peer feedback to (and also receiving peer feedback from) another student, and (3) revising the draft version based on the feedback to construct a final version of the abstract. During phase 1, writing a draft version, students had to select essential content from the article and process this information into an abstract. This abstract was written on a student's individual wiki page. All individual wiki pages of the group members were linked to each other through the overview page of the wiki. For the second phase, students were assigned to provide and eventually receive peer feedback on the draft version, of one particular group member. The peer feedback process was however not reciprocal to avoid influences of received feedback on the feedback given in cycle 2 and 3. This means that students (e.g. student A) received feedback from one specific peer during the complete task (the same one for all the cycles, e.g. student B) and provided feedback to another peer during all three cycles (but again three times the same one, e.g. student C). In order to provide peer feedback, students were required to read the peer's article and formulate their peer feedback regarding the peer's draft on a particular provided template, depending on the condition (see later). This peer feedback template was to be uploaded in the wiki environment and linked to their peer's wiki page comprising the draft. After receiving peer feedback, the third phase required students to adapt their draft version based on their peers' recommendations and own insight. Students were asked to keep their original draft version, i.e. they had to construct their final version at the bottom of their wiki page, indicating their changes in color, i.c. green when they made adaptations based on the feedback, and blue when they modified their initial product based on their own insights. In this way, each wiki page gives a clear overview of the draft, peer feedback received, and final version of each of the three abstracts of one student. As each group consisted out of maximum 5 students, every group worked with a database of 15 different and original scientific articles.

Conditions

For all conditions, the instructor offered a PFB template, which comprised a list of ten predetermined criteria (intention of research, problem statement, methodology, results, conclusion, limitations, structure, language, deadline, and general judgment). This intervention study followed a quasi-experimental design, in which groups were randomly assigned to a particular condition: the no structure condition (groups = 12, N = 57), the basic structure condition (groups = 13, N = 60), or the elaborated structure condition (groups = 12, N = 59). The no structure condition simply received this list of criteria, but was left freely in providing feedback, while the two other conditions received additional instructions on the template. The basic structure condition received the criteria list and two extra guiding questions ('What do you like about your peers' work?' and 'What would you change in your peers' work?'). The elaborate structure condition received a template, which was structured according the principles of feed up, feedback, and feed forward (Hattie & Timperley, 2007), repeating a bullet-pointed criteria list for each of these three principles (see also Gielen & De Wever, 2015). After providing feed up for each criterion in the list, students need to formulate feedback once again for each criterion and finally, finish with feed forward for each criterion separately.

Hypotheses

Taking into account the students' experience and developmental level, instructors have the possibility to differentiate the level of structure they provide during the PA process (Chapman, 1998). The required level of support may vary across students, as one size doesn't fit all (Gregory & Chapman, 2012). Given the fact that we are working with first year higher education students, we believe that the more structure students receive in the PA process, the higher the quality of the peer feedback content eventually will be. Therefore, the following hypotheses are proposed for this study. Students, who receive more structure in their peer feedback template, are more likely to provide peer feedback with:

> **(H1)** a significant higher proportion of elaborations. According to the literature, elaborations contain more relevant information to assist the assessee, while verifications merely state if something is right or wrong (Hattie & Gan, 2011). A significant increase of elaborations could balance the proportion of verifications and elaborations.
>
> **(H2)** a significant higher proportion of negative verifications. Literature revealed that both positive and negative feedback can increase and decrease motivation and performance (Van Dijk and Kluger, 2004). Since students are more inclined to provide positive feedback, a significant increase of negative verifications could balance the proportion of positive and negative feedback.

(H3) a significant higher proportion of general verifications that are focused on particular criteria, and thus a lower proportion focused on the overall performance and language aspects, as feedback should be "on target, objective, focused, and clear" (Shute, 2008, p. 182).

(H4) a significant higher proportion of suggestive elaborations. While informing elaborations provide more details on why a particular criterion was achieved or not, suggestive elaborations provide more suggestions on how the assessee can improve his future performance, which is related to the feed forward component (eg. Hattie & Timperley, 2007). Therefore, a significant increase in suggestive elaborations could be advantageous for the peer feedback content.

(H5) a significant higher proportion of general elaborations that are focused on particular criteria, and thus a lower proportion focused on the overall performance and language aspects. Since elaborations inform learners about their performance and make suggestions for future improvement, it is beneficial when the peer feedback content is focused on particular criteria, instead of on the whole (Shute, 2008).

More specifically, for each of the five hypotheses above, we expect the proportions to be higher for: (Hx.1) the basic structure condition compared to the no structure condition, (Hx.2) the elaborate structure condition compared to the no structure condition, and (Hx.3) the basic structure condition compared to the elaborate structure condition. Stepwise, these effects will be respectively investigated on the actual means (Hx.x.a), and on the means after taking the number of segments into account (Hx.x.b), as we will show later on that there is a large difference between the number of segments in the different conditions. Table 1 presents an overview of all the specific hypotheses.

Table 1

Overview of hypotheses

Hypotheses	*Basic > No (Hx.1)*		*Elab. > No (Hx.2)*		*Elab. > Basic (Hx.3)*	
	Hx.1.a	Hx.1.b (Segments)	Hx.2.a	Hx.2.b (Segments)	Hx.3.a	Hx.3.b (Segments)
H1 – Elaborations	H1.1.a	H1.1.b	H1.2.a	H1.2.b	H1.3.a	H1.3.b
H2 – Negative verifications	H2.1.a	H2.1.b	H2.2.a	H2.2.b	H2.3.a	H2.3.b
H3 – General verifications focused on particular criteria	H3.1.a	H3.1.b	H3.2.a	H3.2.b	H3.3.a	H3.3.b
H4 – Suggestive elaborations	H4.1.a	H4.1.b	H4.2.a	H4.2.b	H4.3.a	H4.3.b
H5 – General elaborations focused on particular criteria	H5.1.a	H5.1.b	H5.2.a	H5.2.b	H5.3.a	H5.3.b

Content analysis

To analyse the content of the peer feedback, a random subsample of nine groups (three groups from each condition) was selected. All three feedback cycles were analysed, which resulted in 123 peer feedback forms from 41 students in total. After the segmentation and coding process, the 123 peer feedback forms resulted in a database of 4717 segments for content analysis. De Wever, Schellens, Valcke, and Van Keer (2006) argued that three aspects are important when conducting content analysis: (1) the choice of the unit of analysis, (2) the choice of the coding scheme, and (3) reporting the interrater reliability of the coding procedure. In the next sections, these issues are shortly detailed.

Unit of analysis

Although the unit of analysis has an important influence on the research focus and coding accuracy, previous studies often neglected to justify their chosen unit of analysis (De Wever, et al., 2006). The unit of analysis defines how the peer feedback content will be divided into fragments, which eventually can be categorised into the content analysis scheme. Following Strijbos, Martens, Prins, and Jochems (2006), a procedural distinction was made between the segmentation and coding process. Firstly, the messages were divided into segments based on the segmentation procedure of Strijbos et al. (2006). Our choice to work with segments as well, can be explained by the fact that we are particularly interested in the detailed and specified content of the peer feedback messages, by focusing on the feedback style, type and focus of each segment. As "sentences or parts of compound sentences will more likely contain a single concept, expression or statement" (Strijbos, et al., 2006, p. 37), we deliberately opted to use the syntactical unit or sentence level (see also Rourke, Anderson, Garrison, & Archer, 2001).

The content analysis coding scheme

As an overarching ready-to-use content analysis scheme fitting our needs did not exist, we developed a coding scheme for analysing the content of peer feedback messages based on a recently developed coding scheme, by Strijbos, et al. (2012), which was in turn based on the generally accepted feedback framework developed by Narciss (2008). Our newly developed scheme includes two categories from Strijbos, et al. (2012), namely the categories 'feedback style' and 'verification type'. Regarding 'elaboration type', the our new coding scheme makes a distinction between informative and suggestive elaborations, referring to providing peer feedback on past performance and providing suggestions in function of future improvement (Hattie & Timperley, 2007). Segments are categorised as informative elaborations, when the feedback on a peer's performance relates to informing, judging, confirming, justifying, etc. On the other hand, when segments have the purpose to suggest, activate, advise, enable, etc. in function of future performance, they are categorised as suggestive elaborations. Finally, for both

verifications and elaborations, the coding scheme takes into account the focus of the PFB segment by paying attention if a particular segment gives general or specific details on the overall assignment, on particular criteria or on language features. To summarize, the newly developed coding scheme attempts to identify variations in the PFB content quality, by concentrating on the peer feedback style, type, and focus. Table 2 represents the five coding categories: peer feedback style, verification type and focus, and elaboration type and focus.

Reliability analysis

For the *segmentation process*, one coder received a training of 4 hours by the researcher, which consisted out of two parts. In the theoretical part, the rules and exceptions of the segmentation procedure (Strijbos, et al., 2006) were openly discussed. Secondly, the practical part involved an initial coding session in which random feedback messages were selected and segmented to familiarise both the coder and researcher with the segmentation procedure. After this training, the peer feedback messages were independently segmented by these two coders. The results showed a proportion agreement of .98, or 390 out of 403 segments were equally segmented. For the *coding process*, one coder (the same one) received a training of 4 hours by the researcher, which consisted out of two parts. In the theoretical part, the scheme was explained with numerous example segments. In the practical part, the coder and researcher coded separately for 2 hours the segments from the previous phase. Afterwards they discussed openly their coding strategy. In order to test the interrater reliability, the coder and researcher next coded separately the feedback messages of one group of each condition, leading to a total double coding of 1506 segments. For the PFB style category Cohen's Kappa was .91, for the verification type category .93, for the verification focus category .94, for the elaboration type category .91, and finally for elaboration focus category .90. As all Kappa values were above the popular benchmark of .80 (Landis & Koch, 1977), there was a high agreement for all categories.

Data analysis strategy

For all categories of the content analysis scheme, analyses of variance were performed to compare the effect of the three different interventions, to increase participants' peer feedback content quality. In the first phase, we performed ANOVAs with the type of condition (no structure, basic structure and elaborate structure) as independent variable, and the proportion scores of the different categories as dependent variables. In order to take into account the number of segments, we ran ANCOVA's with the same independent and dependent variables, but we added the number of segments per student as covariate. As only two possible answers were taken into account for the categories peer feedback style, verification type and elaboration type, we deliberately opted for ANCOVA's on the proportion scores for analysing these binary variables, instead of binary logistic regression (Agresti, 2002). As there were four possible answers for the categories verification focus and elaboration focus, MANCOVA's were used to analyse the data, applying a Bonferroni correction.

Table 2

Coding scheme for analysing peer feedback content

Category	Subcategory	Description	Examples
Peer feedback style	Verification	Is the feedback sentence an evaluative statement expressed as a positive or negative remark on past performance, based on an initial criteria or not?	*Your limitations are not included in the abstract.* *Well written!*
	Elaboration	Is the feedback sentence an informative statement that builds further on verification or remark expressed as e.g. a question, a confirmation, a suggestion or a justification?	*Your limitations are lacking, so please try to include them in your final version.* *I like it because you use your own words.*
	General	Is the feedback sentence a neutral statement, which doesn't have the characteristics of a verification or elaboration?	*This week, I'm providing feedback on your second abstract.*
Verification type	Positive	Is the feedback sentence a positive evaluative statement?	*The intention of the study is well formulated!*
	Negative	Is the feedback sentence a negative evaluative statement?	*I can't find your limitations in the draft!*
	Neutral	Is the feedback sentence a neutral evaluative statement?	*In your abstract, you refer to the methodology.*
Verification focus	Abstract general	Is the feedback sentence an evaluative statement that gives general details about the overall abstract, but without referring to particular criteria?	*All necessary components are included in your draft version.*
	Criteria general	Is the feedback sentence a general evaluative statement that provides minimal details about a particular criteria, or that merely expresses if a particular criteria is correct, present, or not?	*The problem statement and research purpose are present*
	Criteria specific	Is the feedback sentence an evaluative statement that provides profound specific details about the extent to which particular criteria were met in the past performance?	*The introduction summarises perfectly the intention of the research, by mentioning the research purpose before stating the actual context of the research.*
	Language	Is the feedback sentence an evaluative statement about language features such as verbs, translations, pronouns, spelling, grammar, sentence construction and layout?	*There are some little spelling mistakes in your conclusion.*
Elaboration type	Informative	Is the feedback sentence an informative statement, which gives more details about a previous evaluative statement without activating the student to adapt his work?	*Your intro is well formulated! (Pos. Verification) ... Particularly, I like how your abstract deals with the shift from the intention of the study towards the problem statement.*
	Suggestive	Is the feedback sentence a suggestive statement, which gives more details about a previous evaluative statement with the purpose to activating the student to adapt his work?	*In your final version, you should integrate the limitations, which you can find on page 9.*
Elaboration focus	Abstract general	Is the feedback sentence an elaboration that gives general details about the overall abstract, but without referring to particular criteria?	*I believe you can still improve the quality of your abstract*
	Criteria general	Is the feedback sentence a general elaboration that provides minimal details about a particular criteria, or that merely expresses if a particular criteria is correct, present, or not?	*Maybe you should try to merge more the intention of the research and the problem statement*
	Criteria specific	Is the feedback sentence an elaborated that provides profound specific details about the extent to which particular criteria were met in the past performance?	*I would add the number of participants and more details about the context in the methodology section*
	Language	Is the feedback sentence an evaluative statement about language features such as verbs, translations, pronouns, spelling, grammar, sentence construction and layout?	*Once you finish, please check for spelling mistakes*

Results

Descriptives

The descriptives show significant differences between the conditions regarding the number of segments per student [F (2,38)=67.149, p<.001, partial eta squared=.78]. In more detail, the descriptives revealed that students from the no structure (n=14) and basic structure condition (n=14) had respectively 1004 and 1067 segments in total, while the elaborate structure condition (n=13) had 2646 segments in total in their peer feedback messages. Consequently, the elaborate structure condition (M=203.54, SD=69.17) had a significant higher number of segments per student compared to the no structure (M=71.71, SD=25.22), p<.001 and the basic structure condition (M=76.21, SD=27.89), p<.001. For this reason, we calculated the adjusted proportion for each of the five categories, by taking into account the number of segments.

Peer feedback style: Verification or Elaboration

For hypothesis 1, a one-way analysis of variance revealed a significant main effect of condition [F (2, 38)=8.6, p=.001, partial eta squared=.31]. The proportion of elaborations for the basic structure group (p=.001, confirming H1.1.a) and the elaborated structure group (p=.001, confirming H1.2.a) was significantly lower compared to the no structure group. No significant differences were found between the basic and elaborate structure condition (p=.761, not confirming H1.3.a). Taking into account the number of segments, a one-way analysis of covariance revealed a significant main effect of condition [F (2, 37)=5.7, p=.007, partial eta squared=.23]. As shown in Table 3, only the basic structure condition was significantly different from the no structure condition (p=.002 confirming H1.1.b). Although the elaborate structure group provides a higher proportion than the no structure condition, no significant differences were found (p=.349, not confirming H1.2.b). Finally, results also revealed no significant differences between the basic and elaborate structure condition (p=.434, not confirming H1.3.b).

Table 3

Peer feedback style: Descriptives, mean proportion of elaborations per student, and adjusted proportions using number of segments per student as a covariate

	No structure		*Basic structure*		*Elaborate structure*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
M elaborations / student	26.86	3.08	35.86	3.79	105.69	8.21
Mean Proportion elaborations	.457 $^{x, y}$	.109	.573 ^{x}	.084	.584 ^{y}	0.66
Adjusted Proportion (segments)	.478 ^{z}	.112	.592 ^{z}	.108	.573	.162

Note: same superscripts $^{x, y, z}$ indicate significant differences at $p < .01$

Verification type: Positive or Negative

To answer hypothesis 2, a one-way analysis of variance showed no significant main effect of condition [F (2, 38)=.104, p=.901, partial eta squared=.005, not supporting H2.1.a, H2.2.a and H2.3.a] regarding the mean proportion of negative verifications. Taking into account the number of segments, a one-way analysis of covariance indicated a significant main effect for the condition, [F (2, 35)=3.65, p=.036, partial eta squared=.17], and a significant interaction effect between the condition and the number of segments per student, [F (2, 35)=3.41, p=.044, partial eta squared=.16]. The adjusted proportion of negative verifications for the basic structure group was only marginally significant higher compared to the no structure condition (p=.055, nearly to confirming H2.1.b). Between the elaborate structure and no structure group, results revealed no significant difference (.893 not confirming H2.2.b). Additionally, results revealed that the elaborate structure condition had a significant higher proportion of negative verifications compared to the basic structure group (p=.015, confirming H2.3.b). Although the differences between the proportions do not seem to be large, they are significant when controlled for the number of segments.

Table 4

Verification type: Descriptives, mean proportion of negative verifications per student, and adjusted proportions using number of segments per student as a covariate

	No structure		*Basic structure*		*Elaborate structure*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
M Negative verifications / student	4.92	2.64	4.71	4.33	11.69	4.60
Mean Proportion negative verifications	.173	.076	.179	.149	.192	.082
Adjusted Proportion (segments)	.145	.157	.264 [x]	.194	.276 [x]	.194

Note: same superscripts [x] indicate significant differences at $p < .05$

Verification focus

To answer hypothesis 3, A MANOVA using Wilk's statistic indicated a significant main effect for the condition, [F (6, 72)=5.08, p<.001, Wilk's Λ = 0.493, partial eta squared=.29]. Following, separate univariate ANOVAs using a Bonferroni correction, on the outcome variables pointed out significant differences between the three conditions regarding the proportion of verifications that are focused on the overall product [F (2, 38)=6.20, p=.005, partial eta squared=.24] and general verifications that are focused on particular criteria [F (2, 38)=16.09, p<.001, partial eta squared=.46]. Specific verifications focused on particular criteria [F (2,

38)=1.70, p=.196, partial eta squared=.08] and verifications focused on language aspects [F (2, 38)=2.79, p=.074, partial eta squared=.12] appeared not to be significantly different and therefore, these last two are left out in the further analysis. Between the no structure and basic structure group, results revealed no significant difference (p=.168, not confirming H3.1.a). Results indicated that the elaborate structure condition has a significantly higher proportion of general verifications that are focused on particular criteria, compared to the no structure condition (p<.001, confirming H3.2.a) and the basic structure condition (p=.001, confirming H3.3.a). Consequently, the elaborate structure condition has a significantly lower proportion of verifications that are focused on the overall product, compared to the no structure condition (p=.001, confirming H3.2.a) and the basic structure condition (p=.040 confirming H1.3.a). Between the no structure and basic structure group, results revealed no significant difference (p=.123, not confirming H3.1.a).

Table 5

Verification focus: Descriptives, mean proportion of negative verifications per student, and adjusted proportions using number of segments per student as a covariate

	No structure		*Basic structure*		*Elaborate structure*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Verification focus						
General / abstract	124	2.93	81	4.15	147	4.21
General / criteria	172	5.98	180	7.81	600	11.40
Specific / criteria	8	1.15	21	2.59	19	2.93
Language	131	4.94	89	3.05	202	6.14
Mean Proportion						
General / abstract	.298 [u]	.094	.240 [v]	.158	.150 [u, v]	.033
General / criteria	.379 [w]	.104	.448 [x]	.152	.625 [w, x]	.070
Specific / criteria	.017	.031	.047	.074	.016	.029
Language	.304	.124	.264	.124	.207	.052
Adjusted Proportion						
General / abstract	.253	.160	.150	.202	.143	.198
General / criteria	.433 [y,z]	.153	.600 [y]	.190	.657 [z]	.187
Specific / criteria	.021	.074	.085	.093	.001	.093
Language	.293	.157	.165	.198	.201	.198

Note: same superscripts [u, v, w, x, y, z] indicate significant differences at $p < .05$

When taking into account the number of segments, results only indicate that the no structure condition has a significant lower proportion of general verifications that are focused on particular criteria, compared to the basic structure (p=.042, confirming H3.1.b) and the elaborate structure condition (p=.005, confirming H3.2.b). There was no significant difference between the basic and elaborate structure condition (p=1, not confirming H3.3.b).

Elaboration type: Informative or suggestive

After comparing the mean proportion of suggestive elaborations between the three conditions, a one-way analysis of variance showed a nearly significant main effect of condition [F (2, 38)=2.72, p=.079, partial eta squared=.125]. In order to answer hypothesis 4, the proportion of suggestive elaborations for the elaborate structured group was almost significantly higher compared to the no structure group (p=.054, near to confirming H4.2.a) and higher compared to the basic structure group (p=.045, confirming H4.3.a). There was no significant difference between no structure and basic structured group (p=.928, not confirming H4.1.a). Taking into account the number of segments, a one-way analysis of covariance indicated no significant main effect for the condition if we take into account the number of segments as covariate, [F (2, 37)=.119, p=.888, partial eta squared=.006, not supporting H4.1.b, H4.2.b and H4.3.b].

Table 6

Elaboration type: Descriptives, mean proportion of suggestive elaborations per student, and adjusted proportions using number of segments per student as a covariate

	No structure		*Basic structure*		*Elaborate structure*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
M Suggestive verifications / student	224	9.81	297	11.45	947	20.68
Mean Proportion suggestive elaborations	.571 [x]	.184	.566 [y]	.186	.693 [x, y]	.070
Adjusted Proportion (segments)	.603	.209	.587	.183	.636	.288

Note: same superscripts [x, y] indicate significant differences

Elaboration focus

A MANOVA using Wilk's Λ statistic indicated a significant main effect for the condition, [F (6, 72)=5.08, p=.001, Wilk's Λ = 0.540, partial eta squared=.26]. However, separate univariate ANOVAs using a Bonferroni correction, on the outcome variables only indicated significant differences between the three conditions regarding the proportion of general elaborations that are focused on particular criteria [F (2, 38)=11.136, p<.001, partial eta squared=.37]. Other elaborations, which are focused on the overall product [F (2, 38)=2.62, p=.086, partial eta squared=.24], specific elaborations focused on particular criteria [F (2, 38)=1.43, p=.251, partial eta squared=.07] and finally, elaborations focused on language aspects [F (2, 38)=2.55, p=.091, partial eta squared=.12] appeared not to be significantly different between the conditions and therefore, these last three are left out in the further analysis. Similar to the results of verification focus, post hoc comparisons using the LSD test indicated that the elaborate structure condition

has a significantly higher proportion of general elaborations that are focused on particular criteria, compared to the no structure condition (p<.001, confirming H5.2.a) and the basic structure condition (p=.001, supporting H5.3.a). Between the no structure and basic structure group, results revealed no significant difference (p=.551, not supporting H5.1.a).

Table 7

Elaboration focus: Descriptives, mean proportion of elaboration focus per student, and adjusted proportions using number of segments per student as a covariate

	No structure		*Basic structure*		*Elaborate structure*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Elaboration focus						
General / abstract	59	3.11	52	3.85	109	4.31
General / criteria	82	5.06	119	5.05	600	15.05
Specific / criteria	64	4.53	141	9.81	247	15.58
Language	169	7.19	191	7.20	419	12.41
Mean Proportion						
General / abstract	.190	.203	.105	.090	.079	.039
General / criteria	.215 [w]	.163	.245 [x]	.136	.441 [w, x]	.087
Specific / criteria	.157	.134	.249	.198	.171	.116
Language	.437	.163	.399	.182	.307	.096
Adjusted Proportion						
General / abstract	.164	.175	.088	.153	.128	.238
General / criteria	.197 [y]	.179	.234 [z]	.157	.474 [y, z]	.245
Specific / criteria	.218	.198	.290	.175	.062	.270
Language	.422	.205	.389	.179	.336	.281

[u, v, w, x, y, z] Note: same superscripts indicate significant differences

When taking the number of segments into account, a MANCOVA indicated no significant main effect for the condition, [F (6, 70)=1.98, p=.080, Wilk's Λ = 0.731, partial eta squared=.14], and not for the number of segments per student, [F (3, 35)=1.04, p=.384, Wilk's Λ = 0.918, partial eta squared=.08]. Following, separate univariate ANCOVAs using a Bonferroni correction, on the outcome variables reveal some significant differences between the three conditions, but only regarding the proportion of general elaborations that are focused on particular criteria [F (2, 37)=3.78, p=.032, partial eta squared=.17]. Specific elaborations focused on particular criteria [F (2, 37)=2.92, p=.066, partial eta squared=.13] and elaborations focused on language aspects [F (2, 37)=0.30, p=.740, partial eta squared=.01] and finally, elaborations focused on the overall product [F (2, 37)=1.15, p=.327, partial eta squared=.06] appeared not to be significantly different and therefore, these last three are left out in the further analysis. Pairwise comparisons

revealed that students who receive an elaborate structure have a significantly higher proportion of general elaborations focused on particular criteria, compared to the no structure condition (p=.030, supporting H5.2.b) and the basic structure condition (p=.043, supporting H5.3.b), taking into account the number of segments as covariate and using a Bonferroni correction. There was no significant difference between the basic and elaborate structure condition (p=1, not supporting H5.1.b).

As a summary, Table 8 presents an overview of the inspected hypotheses, in which the confirmed ones are highlighted. As the elaborate structure condition showed to have a significantly higher number of segments per student compared to both the no structure and the basic structure condition, we controlled for the number of segments in our analyses. From this point onwards, we will focus on these adjusted proportions to discuss our results. Therefore, the table below presents an overview of the significant differences between the conditions based on these adjusted proportions.

Table 8

Overview of hypotheses taking into account the number of segments

Hypotheses	*Basic > No (Hx.1)*	*Elab. > No (Hx.2)*	*Elab. > Basic (Hx.3)*
	Hx.1.b (Segments)	Hx.2.b (Segments)	Hx.3.b (Segments)
H1 – Elaborations	H1.1.b	H1.2.b	H1.3.b
H2 – Negative verifications	H2.1.b	H2.2.b	H2.3.b
H3 – General verifications focused on particular criteria	H3.1.b	H3.2.b	H3.3.b
H4 – Suggestive elaborations	H4.1.b	H4.2.b	H4.3.b
H5 – General elaborations focused on particular criteria	H5.1.b	H5.2.b	H5.3.b

Notes: (α) Alpha value .05, confirmed hypotheses highlighted, H.2.1.b. is nearly significant with p=.055

Discussion

This study examined how the degree of structuring of a peer feedback template has an impact on the peer feedback content quality. This study attempts to provide more insight into the particular peer feedback content, which students compose with the help of a peer feedback template with a varying structuring degree. In the first part of the discussion, we focus extensively on the proportion of verifications and elaborations (H1) in students' peer feedback messages. In the second part, both the type and focus of verifications (H2 & H3) and of elaborations (H4 & H5) are discussed into detail. Finally, limitations of this study and possible directions for further research are discussed.

With respect to *hypothesis 1*, data analysis indicated that students from all conditions provide peer feedback with a balanced proportion of verifications and elaboration, which can be considered appropriate since previous research claims that successful feedback should include both verifications and elaborations (e.g. Bangert-Drowns et al., 1991; Mason & Bruning, 2001). Furthermore, Hattie and Gan (2011) argued that feedback needs "to move from a predominantly trans missive and verification process to a dialogic and elaborative process in a social context" (p. 257). The results revealed that students in the basic structure condition (59%) have a significantly higher proportion of elaborations, compared to students who receive no structure (48%). The findings also suggest that providing a higher degree of structure in a peer feedback template does not necessarily result in a higher proportion of elaborations. As an elaboration holds the necessary information to assist peers in improving their performance (Hattie & Gan, 2011), this finding implies that adding few guiding questions such as 'What do you like about your peer' work?' or 'What would you change?' increases significantly the elaboration proportion in peer feedback messages, which is in turn beneficial for the peer feedback content quality, as literature states that elaborate and specific feedback results in better performance (Topping, 2010).

As previous literature emphasised that effective feedback quality is determined by both verifications and elaborations (Kulhavy & Stock, 1989; Narciss, 2008), the second part of the discussion gives more details on the particular type and focus of the verifications and elaborations. Related to *hypothesis 2*, the results showed that only students who received an elaborate structure in their peer feedback template, appeared to have a significantly higher proportion of negative verifications, compared to the basic structure condition. However, both basic (26%) and elaborate structure (28%) condition resulted in almost double the proportion of negative verifications compared to students who received no additional structure (14%). This finding suggests that students provide habitually positive verifications and that they are more inclined to provide more negative feedback, when they receive more structure in a peer feedback template. Without neglecting the importance of positive feedback, we believe it is important that students are challenged to formulate negative feedback as well, as this may result in increased effort (e.g. Bandura & Cervone, 1986) and may be needed to point at shortcomings

in students' work. Still, both positive and negative feedback can either increase or decrease performance (Kluger & DeNisi, 1996).

Regarding *hypothesis 3*, this study examined the verification focus by comparing the proportion of general and specific verifications focused on the overall assignment, particular criteria and language aspects. After controlling for the number of segments, both the basic structure (60%) and elaborate structure condition (66%) had a significantly higher proportion of general verifications focused on particular criteria, compared to students who receive no structure (43%). These results suggest that when students receive more structure in their peer feedback template, they provide more general feedback on particular criteria, in which they evaluate if a peers' performance corresponds with the expectations of these particular criteria (Hattie & Gan, 2011). As research of Kluger and DeNisi (1996) claims that feedback should address particular aspects of the task, we believe that peer feedback in which particular criteria are being tackled, is beneficial for the peer feedback process.

Subsequently, the results of elaboration type and focus are being discussed. Inspired by a large body of research (Shute, 2008; Strijbos, et al., 2012; Kulhavy & Wager, 1993; Van den Berg, et al., 2003; Cho & McArthur, 2010; Black & William, 1998; Kluger & Denisi, 1996), this study opted to distinguish elaborations between using informative and suggestive elaborations to formulate an answer on *hypothesis 4*. A large body of research emphasised that feedback should include suggestions for future improvement, and not merely focus on informing students about past performance (e.g. Butler, 1987). In general, students from all conditions provide slightly more suggestive elaborations in their peer feedback messages. This finding is important in view earlier research, which claims that feedback is significantly more effective when it includes details on how to improve, instead of only stating if something is right or wrong (e.g. Bangert-Drowns et al., 1991). All conditions show a similar slight imbalance towards more suggestive elaborations in their peer feedback messages, but there are no significant differences between the no structure (60,3%) basic structure (58,7%) and elaborate structure condition (63,6%). This finding suggests that students provide habitually more suggestive than informative elaborations, and that providing structure in the peer feedback template has no influence on the proportion of informative and suggestive elaborations in peer feedback messages between the conditions.

To answer *hypothesis 5*, this study compared the proportion of general and specific elaboration focused on the overall assignment, particular criteria and language aspects. The elaborate structure condition (47,4%) has a significant higher proportion of general elaborations that focus on particular criteria, compared to the no structure (19,7%) and basic structure (23,4%). Similar to verification focus, these results suggest that when students receive a higher degree of structure in their peer feedback template, they provide more general elaborations that are focused on particular criteria. As feedback content should be usable, focused and well defined (Shute, 2008), feedback that focuses on particular criteria may be more beneficial for the peer feedback content quality than feedback on the overall product or language aspects.

In sum, we can conclude that providing structure in the peer feedback template is a successful instructional intervention for the peer assessment process. This is in line with recent research, which underlines the need for structure and support to ensure effective feedback (Poverjuc, Brook, & Wray, 2012). However, this study also questions if a higher level of structuring necessarily corresponds with higher quality peer feedback. While an earlier study (Gielen & De Wever, submitted) showed that the Feedback Quality Index (adapted from Prins, Sluijsmans, & Kirschner, 2006) scores were significantly higher for the elaborate structure compared to both the no and the basic structure conditions, the present study shows another picture. Results showed that students who receive merely some guiding questions have a higher proportion of elaborations, compared to students without any additional structure, while students in the elaborate structure condition do not necessarily surpass students without any additional structure. Based on previous research (Dillenbourg & Jermann, 2007), students who receive an elaborate structure in their peer feedback template, maybe also be more limited in their creativity and freedom. Taking into account this danger of over-scripting activities (Dillenbourg, 2002), we need to be aware that when students are too heavily structured, this could cause students to provide substantially more peer feedback, which is not necessarily peer feedback of a higher quality.

As a practical implication of this study, we propose for this reason the use of a PFB template for classroom practice, both online and face-to-face, when instructors consider engaging students in PA. This template could include three essential features: a criteria-oriented list, an area to provide feedback, and an area to provide feed forward. First of all, the template needs to provide a list of the pre-specified, or preferably mutual discussed criteria (Sluijsmans, 2002), which have to be considered in order to achieve high quality performance. This criteria list assists the assessor in formulating judgements on particular criteria of a peers' performance, while the assessee receives peer feedback that is focused more individual past performance are being discussed more precisely. Secondly, the template needs to encourage students to provide peer feedback on how well these criteria are achieved in past performance. Finally, the template needs to stimulate students to provide feed forward on how future performance could be improved. In the basic structure condition, the two guiding questions refer respectively to the feedback questions of the framework of Hattie and Timperley (2007), regarding feedback ('What do you like about your peers' work?') and feed forward ('What would you change in your peers' work?'). The latter question guarantees that students receive also feed forward which activates the in function of future performance (Carless, 2007). One way of realizing this may be to provide students with a comprehensive table in which these three essential features are integrated.

Finally, the research findings of this study may also have implications for academics and others who are involved in theory building. First of all, a content analysis scheme has been developed to analyse the feedback content of PFB messages in more depth, which students provide to each other during (computer-supported) collaborative learning activities, and which can be context-independently implemented. Secondly, the findings of this particular study, when implementing this content analysis, reveal that when students provide peer feedback on each

others work, they do not only mention if something is right or wrong to the assessee, but they also equally offer information on why this was right or wrong, in combination with ideas to improve their performance. Additionally, results indicate that students mostly provide rather positive comments, while the elaboration component consists almost equally of informative and suggestive comments. It became clear that all these comments appeared to focus mainly on particular criteria of the performance, instead of solely on the whole assignment or language aspects. As a practical implication of this study, we propose the use of a PFB template for classroom practice, when instructors consider engaging students in PA. This template could include three essential features: a criteria-oriented list, an area to provide feedback, and an area to provide feed forward. First of all, the template needs to provide a list of the pre-specified, or preferably mutually discussed criteria (Sluijsmans, 2002), which have to be considered in order to achieve high quality performance. This criteria list assists the assessor in formulating judgements on particular criteria of a peers' past performance. Secondly, the template needs to encourage students to provide peer feedback on how well these criteria are achieved in past performance. Finally, the template needs to stimulate students to provide feed forward on how future performance could be improved. In the basic structure condition, the two guiding questions respectively refer to the feedback questions of the framework of Hattie and Timperley (2007), regarding feedback ('What do you like about your peers' work?') and feed forward ('What would you change in your peers' work?'). The latter question guarantees that students also receive feed forward, which activates the assessee in function of future performance (Carless, 2007).

Conclusions

In a first-year higher education wiki-based CSCL environment, this study examined the added value of structuring the PA process with the aim to increase the peer feedback content quality, through a collaboration script in which students used a peer feedback template with varying structuring degree. As feedback content is a crucial element for feedback effectiveness, this study investigated in particular the proportional differences of peer feedback content categories between the no structure, basic structure, and elaborate structure conditions.

Regarding *peer feedback style*, the findings pointed out that all conditions provide a reasonably balanced proportion of verifications and elaborations in their peer feedback messages. In more detail, structuring the peer feedback template by adding a few guiding questions, expands the proportion of elaborations significantly in peer feedback messages, compared to students who receive no further structure. Regarding *verification type and focus*, all conditions habitually provide positive and general verifications that are focused on particular criteria. When students receive an elaborate structure in their peer feedback template, they provide significantly more negative verifications than students who receive merely some

guiding questions. When students receive no structure, a majority of the verification segments tend to be positive and focused on the overall assignment and on language aspects. Regarding *elaboration type and focus*, all conditions have slightly more suggestive than informative elaborations, but there was no significant difference between the conditions. The elaborate condition has a significantly higher proportion of general elaborations that are focused on particular criteria, compared to students who receive less structure in the PA process.

One limitation of this study is that the data of only 9 randomly selected groups out of 38 were selected for segmentation and coding. Due to work constraints, it was not feasible to include more groups for the data analysis. Therefore, findings of this study could be expanded and replicated with larger samples, more diverse student populations and a variety of courses. The present study tries to fill in gaps in existing research regarding varying collaboration scripts, in which the PA process is being structured to increase the peer feedback content quality. Furthermore, as starting point for future experimental research, this study provides a content analysis scheme to analyse peer feedback messages in different contexts. Additionally, the study proposes to implement a peer feedback template for the assessor comprising a list of criteria, a feedback and a feed forward component, which combines both the beneficial features of the basic and elaborate structure condition, as a valuable instructional intervention in the PA process to augment students' peer feedback content quality. A final remark could be that this study did not take into account the assessees' evaluation of the received peer feedback, to eventually close the feedback loop (Boud, 2000). Therefore, a suggestion for future research could be including the evaluation of the received peer feedback in the peer feedback template as a fourth element.

The aim of this study was to find out how structuring the PA process, by applying a peer feedback template with a varying structuring degree, can have a beneficial influence on the peer feedback content. Based on the findings of this study, a varying structuring degree in a peer feedback template during the PA process can have an impact on the specific peer feedback content. This study provides some evidence to suggest the use of a structured peer feedback template for peer feedback practices, with the underlying purpose to increase the potential impact of PA and boost students' learning in higher education. This study illustrated how a practical instructional intervention in the feedback process can increase the potential impact of PA and boost students' learning in higher education.

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Wiley, New York.

Anseel, F., & Lievens, F. (2006). Certainty as a moderator of feedback reactions? A test of the strength of the self-verification motive. *Journal of Occupational and Organizational Psychology, 79*, 533-551.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38*, 92-l13.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instruction effect of feedback in test-like events. *Review of Educational Research, 6*, 218-238.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 7-74.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.

Boud, D., & Molloy, E. (2013). Rethinking Models of Feedback for Learning: The Challenge of Design. *Assessment & Evaluation in Higher Education, 38*, 698-712.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*, 474-482.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*, 57-66.

Chapman, E. S. (1998). Key considerations in the design and implementation of effective peer-assisted learning programs. In K. J. Topping, & S. Ehly (Eds.) *Peer-Assisted Learning* (pp. 67-84). Mahwah, NJ: Erlbaum.

Cheng, K.H., Liang J.C., & Tsai, C.C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78-84.

Cheng, W., & Warren, M. (1997). Having second thoughts: Students perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239.

Cho, K., & MacArthur, C. (2010) Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Coll, C., Rochera, M.J., & De Gispert, I. (2014). Supporting online collaborative learning in small groups: teacher feedback on learning content, academic task and social participation. *Computers & Education, 75*, 53-64.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & Education, 46*, 6-28.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010). Roles as structuring tool in online discussion groups: The differential impact of different roles on social knowledge constructions. *Computers in Human Behavior, 26*, 516-523.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen: Open University of the Netherlands.

Dillenbourg P., & Jermann P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (eds.), *Scripting computer-supported collaborative learning: Cognitive, computational, and educational perspectives* (pp. 275-301). New York, NY.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research in computer-supported collaborative learning: from design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3–19). Springer.

Dysthe, O. (2004). The challenges of assessment in a new learning culture. The 32nd International NERA/NFPF Conference, Reykjavik, Iceland.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83*, 70-120.

Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 157-166). London: Kogan Page.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational psychologist, 48*, 56-66.

García, A. S., García-Álvarez, M. T., & Moreno, B. (2014). Analysis of assessment opportunities of learning spaces: On-line versus face-to-face methodologies. *Computers in Human Behavior, 30*, 372-377.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585-594.

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31,* 435-449.

Gregory, G. H., & Chapman, C. (2012). *Differentiated instructional strategies: One size doesn't fit all*. SAGE.

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133-152.

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.

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

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., Hakkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 211-224.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - A conceptual analysis. *Educational Psychology Review, 18*, 159-185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708-721.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*, 279-308.

Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. V. Dempsey, & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 3-20). Englewood Cliffs, NJ: Educational Technology.

Landis, J. R., & Koch, G. G. (1977). A one way components of variance model for categorical data. *Biometrics, 33*, 671-679.

Lockhart, C. & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning, 45*, 605-655.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* NE: Center for Instructional Innovation, University of Nebraska-Lincoln.

Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers In Human Behavior, 27*, 1643-1650.

Narciss, S. (2006). *Informatives Tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse (Informative Tutorial Feedback).* Münster, Waxmann.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner & R. Brunken (Eds.) *Instructional design for multimedia learning* (pp. 181-195). Münster, Waxmann.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science, 37*, 375-401.

O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In: A. M. O'Donnell, & A. King (eds.), *Cognitive perspectives on peer learning* (pp. 179-196). Mahwah, NJ: Erlbaum.

O'Donnell, A. M., & Dansereau, D. F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 120-141). Cambridge, MA: Cambridge University Press.

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

Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior, 26*, 151-161.

Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*, 45-67.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465-477.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Rogers, C. (1951). *Client-centered Therapy: Its Current Practice, Implications and Theory.* London: Constable.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12*, 8-22.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-44.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education, 5*, 77-84.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*, 535-550.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Sluijsmans, D. (2002). *Student involvement in assessment: the training of peer assessment skills*, unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29-48.

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.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages.* Paper presented at the EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

Topping, K. J. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research, 68*, 249–276.

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*, 149-169.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*, 339-343.

Vargas-Vera, M., Nagy M. and Ordóñez de Pablos, P. (2013), "A Framework for Detecting and Removing Knowledge Overlaps in a Collaborative Environment: Case of Study a Computer Configuration Problem", *Journal of Web Engineering, 12*, 422-438.

Van-Dijk, D., Kluger A. N. (2004). Feedback sign effect on motivation: Is it moderated by regulatory focus? *Applied Psychology: An International Review, 53*, 113-135.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. *Assessment and Evaluation in Higher Education, 31*, 9-16.

Van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270-279.

Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education, 40*, 232-247.

Wecker, C., & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: the role of distributed monitoring. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the CSCL 2007* (pp. 763-771). Mahwah, NJ: Erlbaum.

5 Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement.

This chapter is based on:

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education, 88*, 370-386.

Chapter 5
Structuring Peer Assessment: Comparing the Impact of the Degree of Structure on the Peer Feedback Content

Abstract

This study investigates how an instructional intervention focused on engaging both the assessor and assessee in the peer feedback process can be advantageous for the quality of students' peer feedback and written product in a wiki-based computer-supported collaborative learning environment in the first year of higher education. The main aim was to examine the effect of structuring the role of the assessee and/or assessor by respectively providing them with a peer feedback request and/or content checklist, together with a structured peer feedback template. The present study adopted a 2x2 design, in which four conditions were compared: (1) a control condition, (2) a feedback request condition, (3) a content checklist condition, and (4) a combination (feedback request + content checklist) condition. Every student (N=125) belonged to a group (n=27) of five and had to fulfil three consecutive assignments, each consisting of writing an abstract for a scientific paper in the wiki. The results revealed that the quality of both peer feedback and the final product increased for all conditions over time, but no significant differences were found between the conditions at time 2 and time 3. However, when the role of the assessee is structured to request for particular peer feedback, this appeared to be favourable for the peer feedback scores, but only at the initial stage of performance. Building on this, limitations, practical implications, and directions for future research are presented.

Introduction

The role of peer feedback in students' learning

With regard to assessment *for* learning, formative assessment is "specifically intended to provide feedback on performance to improve and accelerate learning" (Sadler, 1998, p. 77). Feedback can be perceived as a practice of formative assessment, which attempts to close the gap between current and desired performance (Sadler, 1989). As an embraced method of formative assessment, PA has been attributed a lot of potential (Black & William, 1998). In this respect, a continuously growing body of research pointed out the value of PA both as an assessment tool (e.g. Cheng & Warren, 1997) and as a learning tool (e.g. Topping, 1998). PA challenges learners in providing feedback on a peer's performance and in receiving feedback

from a peer on one's own performance. However, we cannot assume that all students will be competent to offer high quality feedback for several reasons such as proficiency (eg. Cheng, Liang, and Tsai, 2015). In this respect, previous research emphasised on the fact that students will require unique skills to proficiently perform their role as assessor and assessee (Hovardas, et al., 2014). More specifically, learners develop skills to compile judgments about the quality of a peer's work, based on specific expectations of high-quality work (Topping, 1998). Based on this, the present study focuses on peer feedback as an educational approach of PA.

Following Hattie and Timperley (2007), in order to enhance learning when there is a discrepancy between what is understood and what is aimed to be understood, feedback should provide answers on three major feedback questions: 'Where am I going?', 'How am I going?', and 'Where to next?'. To improve performance, previous research has emphasised on identifying which feedback features should be included or excluded to benefit the understanding of feedback (e.g. Nelson & Schunn, 2008). Feedback content appears to be crucial for the impact of peer feedback on learning and performance (e.g. Cho & MacArthur, 2010). Related to this, earlier research investigated simple versus elaborated feedback (Narciss, 2006; 2008) and concise general versus elaborated specific feedback (Strijbos, et al., 2010). Topping (2010) comments that elaborated and specific feedback leads to better performance. Although a growing body of research claims that feedback has a powerful impact on both learning and performance (e.g. Nelson & Schunn, 2008), a review study revealed recently that more research on the *impact of peer feedback on learning and performance is needed* (eg. Evans, 2013).

In collaborative learning, students work together in small groups to achieve common goals, and in which problems will be tackled more efficiently compared to when students would work individually (eg. Slavin, 1995). Although a peers' competence and lack of objectivity are often mentioned in the literature as examples of possible constraints (eg. Kaufman & Schunn, 2010), research assumes that students acquire more in-group than alone (Dochy, Segers, Van den Bossche, & Gijbels, 2003), which turned out to be also beneficial for students' motivation, social skills, and self-efficacy (Johnson & Johnson, 1994). When students correct the work of other group members and provide feedback in small groups, peer assessment (PA) has proven to be beneficial for the assessor (e.g. Topping, 1998), and assessee (e.g. Tsivitanidou, Zacharia, & Hovardas, 2011), both as an assessment tool and a learning tool (e.g. Evans, 2013). Previous literature on peer assessment emphasises that actively involving students in the assessment process boosts not only their understanding of it (eg. Boud & Molloy, 2013), but that it also increases students' engagement in their own learning (De Wever, et al., 2011). Other research claims that producing feedback reviews "engages students in multiple acts of evaluative judgement, both about the work of peers, and, through a reflective process, about their own work; that it involves them in both invoking and applying criteria to explain those judgements; and that it shifts control of feedback processes into students' hands" (Nicol, Thomson, & Breslin, 2014, p. 102).

When learners are engaged in the role of assessor and assessee, particular PA skills are required. As an assessor, learners need to be able to recognize and assess particular criteria, judging the performance of a peer, and eventually provide peer feedback (PFB). Compared to this, assessees traditionally need to "critically review the PFB they have received, decide which changes are necessary in order to improve their work and proceed with making those changes" (for detailed description see, Hovardas, Tsivitanidou, & Zacharia, 2014, p. 135). PA engages students in cognitive activities such as summarising, explaining, providing feedback, identifying mistakes and gaps, which are dissimilar from the expected performance (Van Lehn, et al., 1995). PFB intends to involve learners by providing and receiving opinions, ideas and suggestions for improvement to their peers (Black & William, 1998).

Previous research claims that online assessment is more beneficial compared to face-to-face assessment (Tsai 2009; Tsai and Liang 2009; Yang and Tsai 2010). A particular reason could be that online technology offers learners more freedom in time and space (Tsai, Lin, & Yuan, 2002). When students are involved in an online PA learning environment, previous research found that students were highly confident and strongly intrinsically motivated (Tseng and Tsai, 2010). In this respect, a growing body of research illustrated that peers' writing performance can be enhanced, when learners are involved learners in online PA activities (eg. Cheng, Liang, & Tsai, 2015; Gielen & De Wever, 2015; Tsai & Chuang, 2013; Xiao & Lucking, 2008). Related to this, online PA appears to be more ideally for augmenting the validity and reliability of PA, when it is integrated into online modules instead of in a traditional setting ((Tsai & Liang, 2009; Mostert & Snowball, 2013)." As wikis are free and easily accessible, in which peers can work in group by contributing, commenting and editing further on each other's work, other research claims wikis are ideally suited for collaborative learning (Ebner, Kickmeier-Rust, & Holzinger, 2008). Being an example of an effective online learning environment (Ertmer, Newby, Yu, Liu, Tomory, et al., 2011), wikis are often praised as a computer-supported collaborative learning (CSCL) tool to support PA activities and online collaboration (De Wever, et al., 2011). According to Sadeghi and Kardan (2015), CSCL "can provide a more efficacious, more convenient and more flexible collaborative learning experience for both learners and instructors" (2015, p. 437). More particular, implementing wiki-tasks in higher education appeared to be beneficial for students' learning (De Wever, et al., 2015).

Although feedback has proven to be advantageous for both learning and performance (e.g. Nelson & Schunn, 2008), it appears that not all feedback automatically results in performance improvement (Kluger & Denisi, 1996). Earlier research shows that providing feedback was more beneficial for the assessor's future performance than that of assessees who simply received feedback (Kim, 2009), as students do not necessarily know what to do with the assessor's feedback (Sadler, 1998). Hattie and Gan (2011) also discussed the unpredictable effect of feedback. The framework of Hattie and Timperley (2007), who describe feedback as the component clarifying how well the assesse achieved particular criteria and feed forward as the component offering suggestions how the assessee can actually improve future performance, may be used to explain this: when feed forward is missing, assessees may not know what to do. Related to this, previous research already indicated that further research on the impact of PFB

on students' learning and performance is needed (e.g. Evans, 2013; Hattie and Timperley, 2007). Inspired by this, the current research attempts to determine *how the roles of assessor and assessee in PA practices should be tailored in order to optimize the PFB process in function of students' learning and performance*.

Scripting the peer feedback process in function of students' learning

Being a complex learning task, PA requires "high-level" cognitive processing (King, 2002). Nevertheless, previous research has revealed that these particular processes do not happen instinctively (Kollar & Fischer, 2010) and that students require instructional support to participate in these high-level collaboration processes (e.g. Weinberger, Stegmann, Fischer, & Mandl, 2007; Cole, 2009). Related to this, previous research underlines the importance of structure and support in order to safeguard effective feedback (Poverjuc, Brook, & Wray, 2012). For this reason, researchers and instructors should investigate what type of support for the assessor and assessee is fundamental to foster high quality feedback (Hovardas, et al., 2014).

In this respect, scripting is suggested as a possible solution (Fischer, et al., 2013). Collaboration scripts provide more details on role assignment and specify activities in small groups in order to stimulate successful collaborative learning activities (e.g. Kollar, Fischer, & Hesse, 2006). Scripting collaboration has been shown to be beneficial for the acquisition of domain-general skills (e.g. Noroozi et al., 2013). However, other research revealed contradictory results regarding the effectiveness of collaboration scripts for domain-specific learning outcomes (Kollar et al., 2014). Therefore, investigating the effectiveness of different scripting techniques in various situations is an important field of study (Kollar & Fischer, 2010). Research revealed that collaboration scripts can actually obstruct learner's knowledge attainment when they become too strict or simply too flexible (eg. Fischer, et al., 2013). Therefore, it is essential to determine the precise scripting level that learners require (Dillenbourg, Järvelä, & Fischer, 2009).

Until now, a growing body of research has experimented with varying instructional interventions to enhance the effectiveness of PFB, for instance, by: organizing a training to improve PFB (Sluijsmans et al., 2002), working with multiple raters instead of one (Cho & Schunn, 2007), offering guiding questions to support the assessor while giving PFB (Gielen & De Wever, 2012), providing sentence openers to encourage interaction between students (Baker & Lund, 1997), or by creating a PFB template with a varying structuring degree to provide feedback and feed forward (Gielen & De Wever, 2015). Nevertheless, the majority of all these instructional interventions are habitually focused on the role of the assessor (Gielen, et al., 2010). It is within this frame that this study is particularly interested in *how collaboration scripts can support the role of both assessor and assessee, in function of students' learning*.

Scripting the role of both assessee and assessor

In the assessment literature, there is not much information on instructional interventions to enhance the effectiveness of PFB, which focus on the role of the *assessee.* Gibbs and Simpson (2004) mentioned two recommendations for so-called two-stage assignments. Firstly, the assessee should formulate a PFB request in order to receive personalised feedback, e.g. based on an 'a priori question form', in which the assessee specifies his PFB request according to several performance criteria. Secondly, the assessee should also have the opportunity to respond to the received PFB in order to close the feedback loop. Previous research pointed out the advantages of using such a 'posteriori reply form' to close the feedback loop, as it inspires students to reflect on the received PFB and on how they applied it for text revision (Boud, 2000).

As this study focuses on both the role of the assessor and assessee, it is interested in how an instructional intervention, by engaging the assessee in the PFB process, can have an impact on the PFB quality. Therefore, we discuss the 'a priori question form' in more detail. Gielen et al. (2010) incorporated an 'a priori question form', often referred to as '*feedback on demand*' or '*feedback request*', together with a peer feedback template, in an assessment process that encouraged the assessor to fulfil the feedback request of the assessee. It appeared that instructional interventions might not have an effect immediately, but they do have in the long term. Related to the feedback request, Gielen et al. (2010) wrote that:

> *''Such an intervention may enhance both ''individual accountability'' and ''positive interdependence'' (Slavin, 1989), and motivate and guide assessors to provide 'responsive' feedback (Webb, 1991). It may also result in more appropriate feedback (Webb, 1991) and promote 'mindful reception' (Bangert-Drowns et al., 1991), that is, make assessees feel more personally addressed and subsequently more inclined to apply the feedback''* (Gielen, et al., 2010, p. 308).

A few other examples of studies that embedded a PFB request can be found in the literature, showing a varying degree of delegating responsibility to students. An example to engage students with feedback was asking students "to specify on their assignment, what they would like feedback on, and giving feedback on nothing else" (Gibbs & Simpson, 2004, p. 24). A similar approach was found elsewhere, requiring students to specify their feedback request in more detail, including for example on which specific aspects they desire feedback (Nicol & MacFarlane-Dick, 2006), or in which specific feedback style (Prins, Sluijsmans, & Kirschner, 2006).

As mentioned earlier, instructional interventions that further specify the role of the assessor in the PFB process, are more widely spread in the assessment literature. Previous research gave an overview of some of these examples such as "coaching by providing hints, prompts, and feedback; modelling the use of cognitive strategies by thinking aloud; presenting cue cards, checklists and process worksheets; asking leading questions; and giving part of a solution" (Van Merriënboer, Kirschner, & Kester, 2003, p. 6). In order to support the assessees' learning, the assessor has a dual task when providing PFB, one that requires high-level cognitive processes.

First of all, the assessor needs to deeply process the performance of the assessee and secondly, the assessor needs to engage in planning and monitoring on how to construct valuable PFB for the assessee (Kollar & Fischer, 2010). When assessors are involved in the PFB process, previous research recommends the use of guidelines and checklists (Topping, 2009), as they assist learners to set task-specific goals (Butler, 2002). When the assessor is scripted to complete a checklist, in which essential content concerning a peers' performance is gathered, as a preparation in function of formulating valuable PFB, this instructional intervention could be a possible approach to augment the PFB quality eventually. Previous research underlines that the content of PFB messages is vital for the effectiveness of the feedback (e.g. Cho & MacArthur, 2010). For this reason, it is crucial to further explore alternative instructional interventions, which further specify the role of the assessor and assessee in the PFB process, to shed more light on what kind of support could be necessary to safeguard high quality PFB (Hovardas, Tsivitanidou, & Zacharia, 2014).

Material and methods

Participants

All participants in the study were first-year bachelor students of an Educational Sciences program (N = 125), enrolled in the course 'Instructional Sciences'. Participants were randomly assigned to groups (n = 27) of about 4 to 5 students to work on a wiki task, in which each student was required to individually write a number of academic abstracts for scientific articles. From a number of interviews, which are not yet analyzed in detail, a subsample of 43 students indicated that the majority (33/43) already had an experience with PA during their primary or secondary education. However, only a minority (3/43) indicated that providing peer feedback was commonly used. One student explained she gained experience during her teacher training for primary education. For some students (10/43), providing feedback appeared to be something new. The interviews revealed that mostly PA was implemented in secondary education during the evaluation of presentations and group products. It appeared that the presentations were evaluated through both oral and written feedback, while judging group product happened mostly through written feedback. In both cases the feedback was rather qualitative approach, in which students were required to specify positive aspect, but also points of improvement. Only some students (9/43) mentioned they were required to qualitatively evaluate their peers with scores. Only a minority (2/43) actually used assessment criteria to provide a peer with scores from 1 till 5 for each criterion, while other participants (7/43) had to justify their evaluation through qualitative feedback. In this study, no tests reading proficiency and computer literacy were administered.

Research design and conditions

Each group member had to consecutively read three different academic journal articles and write one abstract for each of these three provided articles (i.e. they received the paper, but the abstract was left out). Each student received three other articles, so there were 15 different articles in each group. For each abstract, students participated in four phases: (1) writing a draft version of the abstract (2) providing non-reciprocal PFB to (and also receiving PFB from) another group member (the same one for the three iterations), (3) revising the draft version based on the feedback to construct a final version, and (4) evaluating the received PFB. All four phases were part of their curriculum requirements.

During the PFB process, the instructor provided the same structured PFB template to every student. This template consisted of a list of seven criteria (intention of research, problem statement, methodology, results, conclusion, limitations, and general judgment), which was structured in a way that students were encouraged to provide both PFB and feed forward for each criterion separately (Hattie & Timperley, 2007). During the first PFB cycle, only the assessor was instructed to employ this PFB template, while the assessee was not involved in the PFB process. As part of the instructional intervention, this study deliberately introduced additional structuring during the second and third feedback cycle in order to engage both the assessee and assessor in the PFB process.

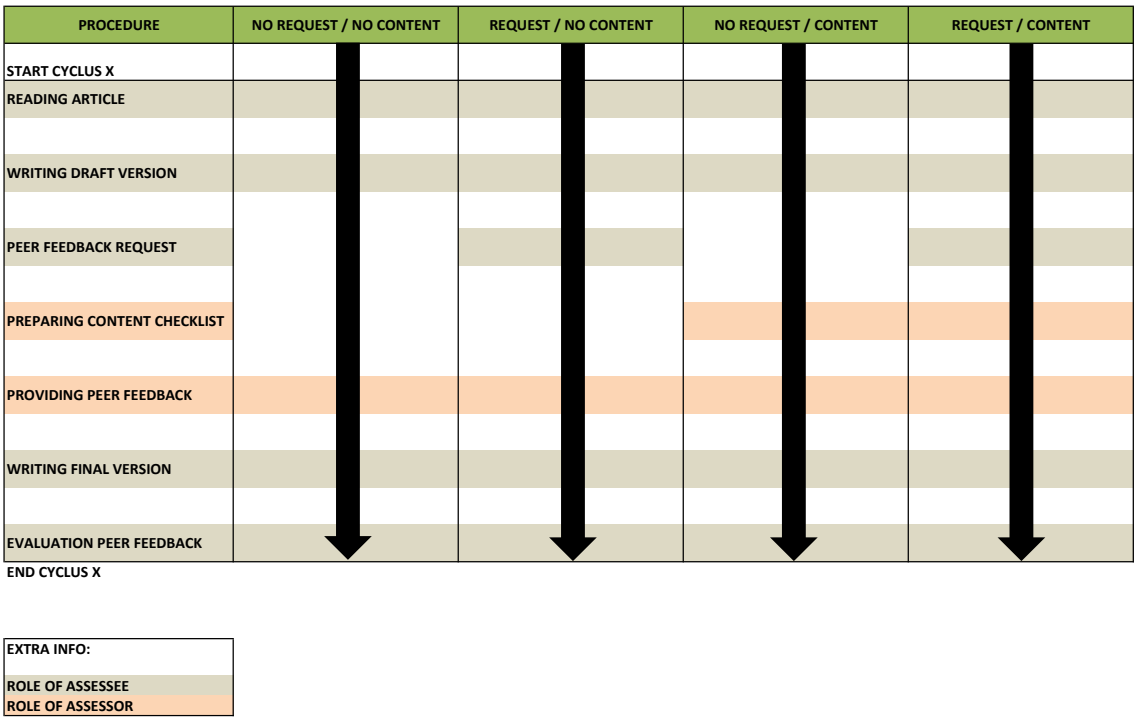


Figure 2. Procedure for each condition

To examine the effects of this script, this study adopted a quasi-experimental 2 x 2 factorial design, which resulted in four conditions: (1) a control condition, in which only the PFB template was provided; (2) a request condition, in which in addition to the PFB template, the assessee was provided with peer feedback request form, which required the assessee to formulate a specific feedback demand; (3) a content condition, in which again the same PFB template was provided, but this time together with a content checklist form, which required the assessor to actually select the essential content from the paper, meaning that the content checklist was contextualized to a specific paper; and (4) a combination (request + content) condition, in which the same PFB template was combined with the peer feedback request form (i.e. structuring the role of the assessee) and a content checklist form (i.e. structuring the role of the assessor).

To sum up, the PFB process in this study consisted out of 4 phases, which are also illustrated in detail in Figure 1: (1) requesting for specific PFB, (2) preparing PFB content, (3) providing feedback on a structured PFB template, based on (a) the PFB request and/or (b) the prepared content checklist, and finally (4) evaluating the received PFB. In the appendix, screenshots provide a visual and chronological overview of all the different steps of one feedback cycle, namely (1) Draft version; (2) PFB request; (3) Content checklist; (4) Formulating PFB; (5) Final version and (6) Evaluating PFB. In order to study the differential impact of the script that further specifies the role of the assessor and assessee during the different PFB phases, this study primarily focuses on the first two PFB phases, in which (1) the assessee requests for particular PFB and (2) the assessor prepares PFB content with the help of a content checklists.

The role of the instructor

In the practical part, the instructor facilitates the learning experience by managing the online learning environment, in which he can for example appoint or randomize students into different groups, decide on certain wiki settings such as privacy, create a group forum, make scientific articles or group documents available for students. First of all, the instructor gave an introduction to this task during a face-to-face lecture of 60 minutes at the start of the academic year. Additionally, the practical website gave an overview of the whole assignment, in which students could find additional information such as FAQ, but also an instructional video, in which all the necessary steps of the wiki assignment are chronologically discussed. During the assignment period, the role of the instructor also requires mentoring and monitoring on a class, group and student level. On the group level, the instructor is required for example to mediate and intervene if particular group members do not respect the deadlines and to monitor if all group members actively participate and contribute to the wiki assignment. On a personal level, the instructor needs to deal for example with students who drop out or for who the assignment seems too overwhelming. As these students are first-year bachelor students, the instructor will offer regularly instructions by posting weekly announcements in the online learning

environment. Students were encouraged to seek help in first instance with students within their group or between other groups. Finally, students could mail their unresolved practical questions to the instructor.

Instruments

Peer feedback request

After finishing the draft version of an abstract, students in the request conditions are required to compile a specific PFB request or 'feedback on demand' (Gielen, et al., 2010), based on the expectations and judgment of their own work. In particular, the assessees have to indicate firstly on which of the seven predetermined criteria they expect feedback and secondly, what kind of specific feedback they expect. When clarifying their specific PFB request on the list of selected criteria, the assessees are left freely regarding how they specifically ask for feedback. This could be done for example by formulating specific questions, or by referring to a particular paragraph, etc. As illustrated in Figure 1, the PFB request has to be submitted before the assessor can start compiling and providing his/her PFB regarding the draft version of the assessee.

Content checklist for the assessor

In order to activate the role of the assessor, students in the content checklist condition are required to complete a checklist, based on the same seven predetermined criteria, so as to be able to deeply process a peers' performance (Kollar & Fischer, 2010). In order to do so, students start off by reading the scientific article of the peer, for who they have to provide PFB in the next phase. Afterwards, students are encouraged to carefully identify the information, which according to the assessor contains the essential components of the scientific article that should be present in the abstract, to meet the expectations of good performance. Next, students are required to sum up and categorise the selected content for all criteria separately. The assessor is scripted to complete this content checklist before actually formulating PFB, as its purpose is to serve as an input source during the third phase in the PFB process.

Structured peer feedback template

For the third phase in the PFB process, the instructor provided all four conditions with an identically structured PFB template, consisting out of four sections: (1) a list of predetermined criteria (2) a section to provide feedback on each criterion, (3) a section to provide feed forward on each criterion and finally, (4) a section to evaluate the received PFB for each criterion. Following, these four sections will be discussed in further detail.

The first section provides a list of seven predetermined criteria, presented in a table. In the second and third section, students were required to provide feedback and feed forward: "How am I going?" (What progress is being made toward the goal?), and "Where to next?" (What activities need to be undertaken to make future improvement?) (Hattie & Timperley, 2007, p. 86). While providing PFB on the structured peer feedback template, the assessor needs to check whether the assessee previously submitted a specific peer feedback request, and, if so, provide specific feedback and feed forward exclusively focussing on the requested criteria, instead of providing feedback and feed forward related to all of the criteria. In addition, the assessors of the content checklist conditions had to actively employ the information of the checklist, when formulating their feedback on the structured peer feedback template. In the fourth section, students were instructed to evaluate their received PFB. The assessees could indicate, for each criterion and its corresponding feedback, whether the received PFB was valuable or rather irrelevant.

Scoring rubric for quality of peer feedback messages (Peer feedback score)

The Feedback Quality Index (Prins et al., 2006) was used to measure the quality of peer feedback messages. Building on this, a series of previous studies (Gielen & De Wever, 2012; 2015) and the present study consistently incorporated this scoring rubric, maintaining all three main categories (use of criteria, nature of the feedback, and writing style), and their sub-categories with corresponding scoring percentages (Prins et al., 2006). This resulted in a scoring rubric of 9 items with a scoring range between 0 and 100 to measure the quality of PFB messages of first-year higher education students. (For more info, see Gielen & De Wever, 2015).

Scoring rubric for quality of the wiki product (Product score)

This study applied a scoring rubric, used previously in two studies (Gielen & De Wever, 2012; 2015), in which the necessary components of a good abstract are integrated. This scoring rubric counted four main categories (situating the study, content of the abstract, style, and general impression) and nine corresponding sub categories. (For more info, see Gielen & De Wever, 2015). Similar to the previous rubric, this scoring rubric for analysing the quality of the wiki product also had a scoring range between 0 and 100.

Students' perception towards PA at the end of the assignment

At the end of the wiki-assignment, students were asked to fill out a questionnaire. Items were measured using 5-point Likert scales, from 1 (totally disagree) to 5 (totally agree). The first main section 'The role of the assessor' consisted of 'Perception towards providing feedback' (e.g., "I believe it is disadvantageous that the assessment happens in an online environment") (7 items), 'Perception towards the PFB request' (e.g., "I believe the PFB request limits me as an assessor

because habitually I would provide more PFB") (7 items) and 'Perception towards the content checklist' (e.g., "I believe the content checklist is useful to increase the quality of my PFB") (5 items). The second main section 'The role of the assessee' attempts to shed more light on the perception of students when they act as assessees. The questions dealt with: 'Perception towards receiving feedback' (e.g., "I believe that the value of the received feedback from an assessor or instructor is equal") (5 items), 'Perception towards the PFB request' (e.g., " I believe the PFB request is an added value for the whole process") (7 items) and 'Evaluation of the received feedback' (e.g., "I believe that generally the quality of the received feedback is insufficiently") (5 items). For all the different items, analyses of variance were performed to compare the effect on students' perception in the four conditions.

Hypotheses

The following hypotheses are put forward with respect to the PFB quality scores:

Since at time 1 no additional forms are introduced and all conditions are only using the PFB template, (H1) we expect the PFB quality scores of students in all four conditions not to differ significantly at the start of performance. With respect to the quality of the PFB over time, (H2A) we expect the PFB quality scores to increase significantly for all four conditions from time 1 to time 3, as students become more proficient assessors when they are involved in multiple practice occasions (e.g. Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). Thus, we also expect a significant increase of PFB scores (H2B) from time 1 to time 2 and (H2C) from time 2 to time 3, for all conditions. As the intervention was introduced during the second iteration, we expect the PFB quality scores of students in the request, content, and combined condition to be higher than those of students in the control condition (H3A) at time 2 and (H3B) at time 3, since previous research has shown that a certain degree of structure can have a beneficial effect on the PA process (Gielen & De Wever, 2015). More specifically, the peer feedback request form is supposed to increase the PFB scores significantly, as it motivates and directs the assessor to formulate more 'responsive' feedback-on-demand (Webb, 1991). As previous research recommended the use of guidelines and checklists when assessors are involved in the PFB process (Topping, 2009), the content checklist is supposed to increase the PFB scores significantly, as assessors are more profoundly prepared and encouraged to formulate high quality PFB content, which is in turn essential for the effectiveness of feedback (e.g. Cho & MacArthur, 2010). With respect to the combined request + content condition, no specific hypothesis is formulated, as there was no literature available that has investigated this combination.

Similar hypotheses are formulated with respect to the product scores. Since at time 1 no additional forms are introduced, and all conditions are only using the PFB template, we also expect (H4) the product scores of students in the four conditions to be the same at the start of performance. With respect to the quality of the writing product over time, we expect (H5A) a

significant increase of the product scores for all four conditions from time 1 to time 3, as practice results in performance improvement over time (eg. Gielen & De Wever, 2015). Thus, we expect that results increase significantly (H5B) from time 1 to time 2 and (H5C) from time 2 to time 3. When students receive more high quality feedback, we assume that this will have a stronger impact on the final writing product, as other research claims that more specific and elaborated feedback stimulates better performance and outcomes (eg. Strijbos, et al., 2010). As the intervention was first introduced in the second iteration, we expect the product scores of students in the request, content, and combined condition to be higher than those of students in the control condition (H3A) at time 2 and (H3B) at time 3. More specifically, the peer feedback request form is supposed to increase the product scores significantly, since previous research has revealed that assessees pay more attention on feedback-on-demand (Gielen, et al., 2010) and response - specific feedback augments learning efficiency (Hansen & Almond, 2007). Following, the content checklist is supposed to increase the product scores significantly, as we believe that high quality PFB content has the potential to improve the writing product more compared to low quality PFB content. However, we need to take into account that all kind of feedback can be beneficial for students' performance (Topping, 1998). With respect to the combined request + content condition, no specific hypotheses is formulated for the combination.

Data analysis

Given the clear hierarchical structure of the data, namely three measurement occasions (i.e. the PFB moments, indicated by the variable 'time' (level 1) are nested within each of the 125 students (level 2), who are in turn nested within 27 groups (level 3), multilevel modelling for repeated measures (MLwiN 2.29) was used to analyse the PFB quality and the product quality (ie. the quality of the versions of the abstract written in the wiki). Firstly, a fully unconditional null model was tested for both PFB score and product. Following this, a compound symmetry model was created by adding the categorical predictor 'time' to the null model, since it is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999). In this model, time 1 is taken as the reference category. Then, the categorical predictor 'condition' is added in the next step for both the PFB and product score. In a final phase, the interaction condition*time was added to the model. By using a stepwise multilevel approach, we could check the additional value of each subset of variables to the model.

Results

Peer feedback score

The null model shows that 24.98% of the total PFB variance is situated at the group level (p=.011), the proportion of variance due to differences between students within groups was 35.81% (p<.001), and finally 39.21% of the total variance is situated at the time level (p<.001; see also Table 1 in appendix). Related to hypothesis 1, results revealed that at the start (i.e. time 1, when the structuring intervention had not yet started), the request condition has significant lower PFB scores compared to the content condition (χ2=7.192, df=1, p=.008) and compared to the combination condition (χ2=6.326, df=1, p=.012), which is in contrast to H1.

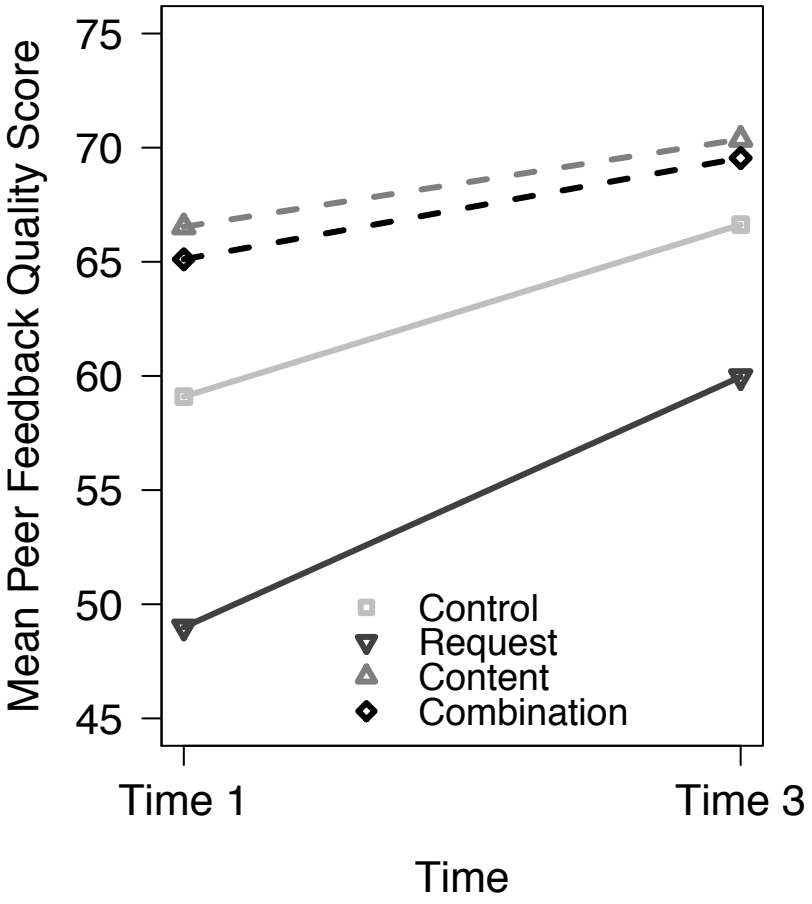


Figure 2A. Evolution of peer feedback scores over time

Figure 2A shows that the PFB scores increase from time 1 to time 3 for all four conditions. When taking a detailed look at the separate conditions, only the scores of the control (χ2=6.385, df=1, p=.011) and request condition (χ2=13.969, df=1, p<.001) increased significantly from time 1 to time 3, which only partly confirms hypothesis 2A. As shown in Figure 2B, the PFB scores increased significantly from time 1 to time 2 for the request condition with 11.88 (χ2=16.381, df=1, p<.001), and with 7.50 for the combination condition (χ2=5.936, df=1, p=.015), which only partly confirms hypothesis 2B. When taking a closer look at the increase in PFB scores from time 2 to time 3, results only show a significant increase for the control condition with 5.94

(χ2=3.969, df=1, p=.046), which only partly confirms hypothesis 2C. When taking a closer look at the PFB scores at time 2 and time 3, no significant differences were found between the four research conditions, not confirming hypothesis 3A and 3B.

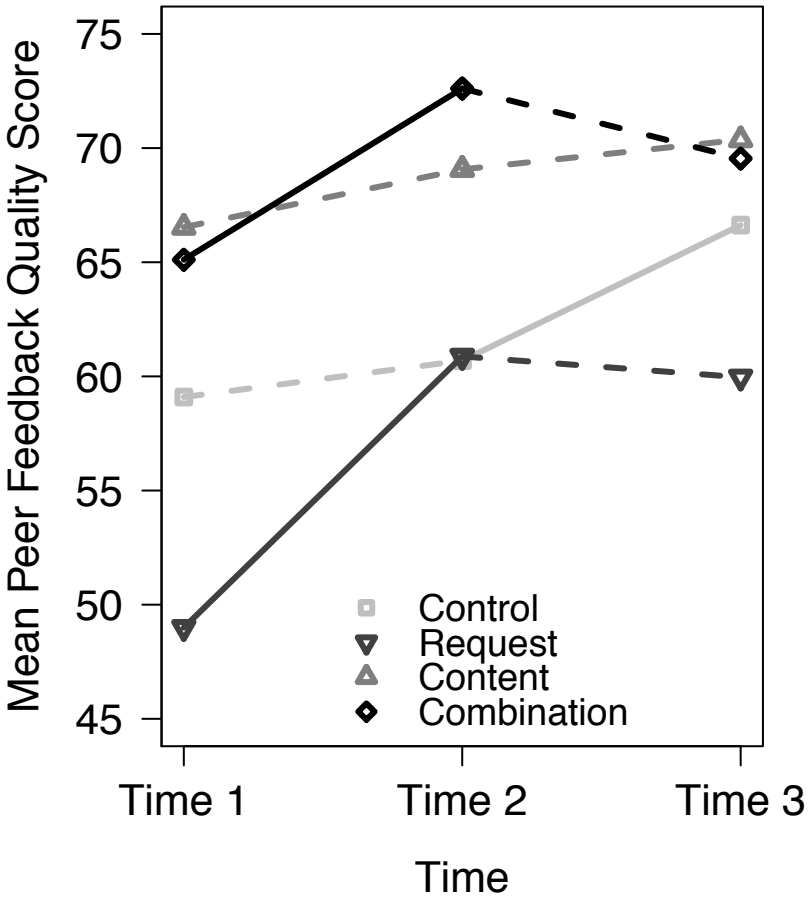


Figure 2B. Peer feedback scores at time 1, time 2 and time 3

Product score

As shown in Table 2 in the appendix, the null model reveals that 1.28% of the total variance of product scores is situated at the group level (p=.684), the proportion of variance due to differences between students within groups was 15.36% (p=.021), and finally 83.36% of the total variance is situated at the time level (p<.001). Looking at hypothesis 1, results revealed no significant differences in product scores between the four conditions at the start (i.e. time 1, when the structuring intervention had not yet started), which confirms hypothesis 4.

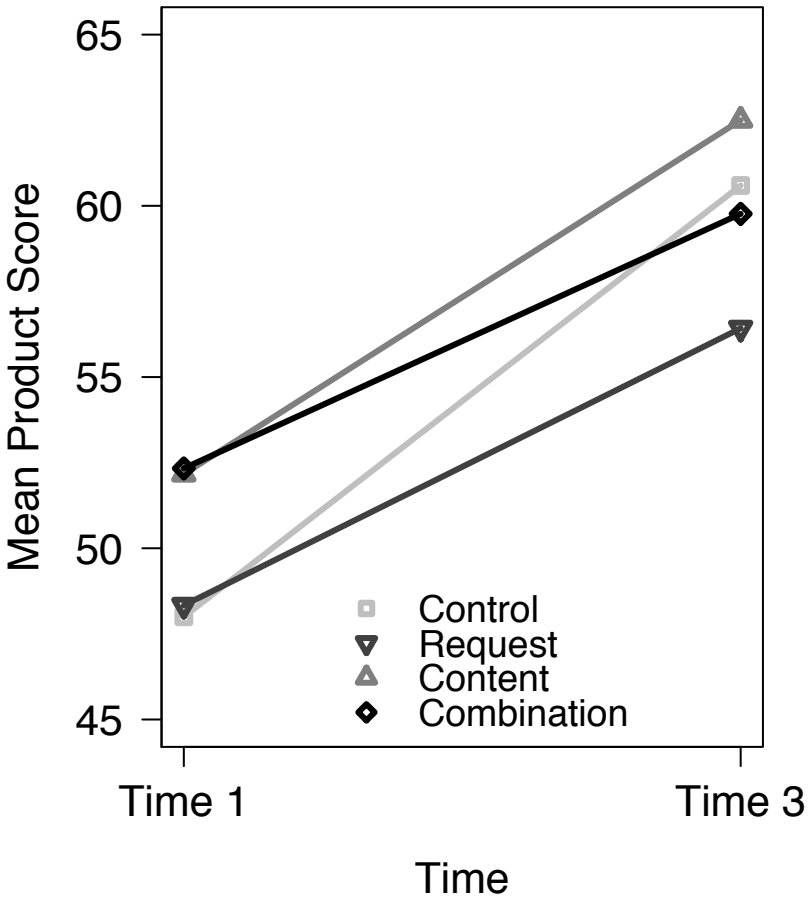


Figure 3A. Evolution of peer feedback scores over time

Figure 3A shows that all conditions improve significantly over time. In more detail, the product scores increased significantly for the control (χ2=18.484, df=1, p<.001), request (χ2=7.868, df=1, p=.005), content (χ2=11.667, df=1, p<.001), and combination condition (χ2=6.037, df=1, p=.014) from time 1 to time 3, which confirms hypothesis 5A. When decomposing the increase over time, results revealed differences between the four conditions regarding the increase of product scores over time. As shown in Figure 3B, the product scores of

three out of four conditions increased significantly after the structuring intervention, which confirms hypothesis 2B, except for the content condition. More specifically, the conditions for who the product scores increased significantly from time 1 to time 2 were the control condition with 9.06 ($\chi2$=9.572, df=1, p=.002), the request condition with 6.82 ($\chi2$=5.587, df=1, p=.018), and the combination condition with 8.70 ($\chi2$=8.270, df=1, p<.004). The content condition increased with 5.40 from time 1 to time 2, which is close to significant ($\chi2$=3.186, df=1, p=.074). Although three out of four conditions had higher product scores at time 3, no significant increase was found for any condition from time 2 to time 3, which does not confirm hypothesis 5C. When taking a closer look at the product scores at time 2 and time 3, no significant differences were found between the four research conditions, contrary to hypothesis 6A and 6B.

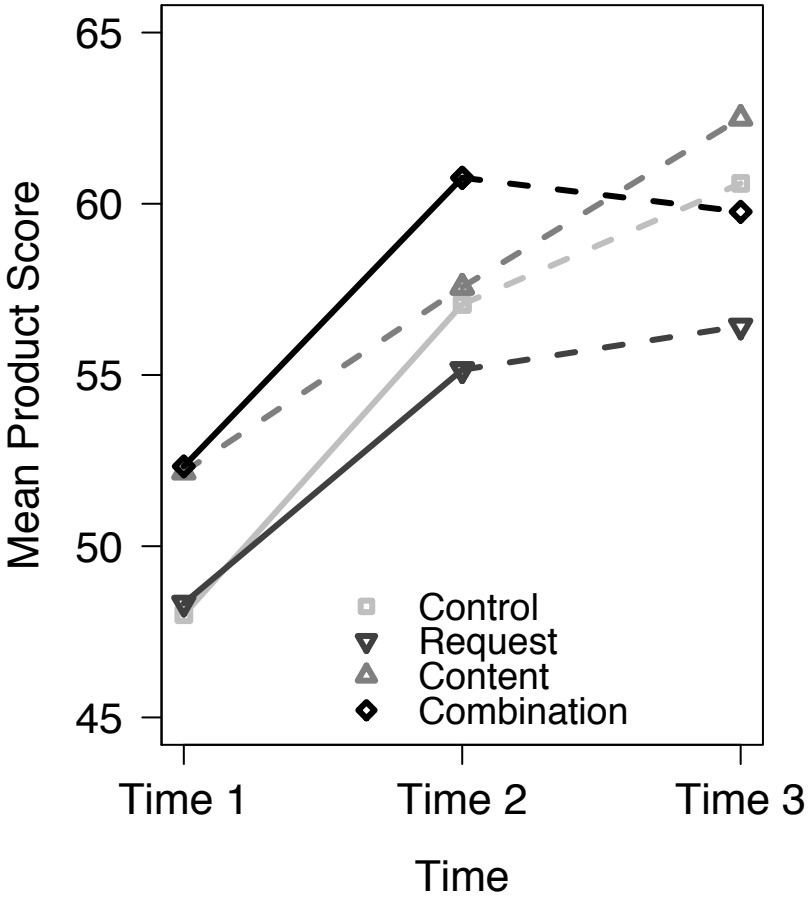


Figure 3B. Product scores at time 1, time 2 and time 3

Students' perception at the end of the wiki-task

Students' perception towards PA

In order to provide some extra background information, a questionnaire revealed that mostly all students had gained some experience with PA before arriving at university (M= 3.07, SD=1.24). In general, providing feedback was not perceived as a difficult task (M= 2.46, SD=0.92). Also, students not perceive it as a shortcoming when the PA process happened in an online setting (M= 2.38, SD=1.08). When evaluating their received PFB, students indicated that they did not perceive the quality as insufficient (M= 2.24, SD=0.98) and they are more or less convinced that the quality of their final version enhanced significantly, thanks to the received feedback (M= 3.50, SD=0.92).

Students' perception towards the additional structure in PFB process

In general, students more or less agreed that the PFB request was useful for formulating PFB in their role as assessor (M= 3.61, SD=0.97). Further analysis revealed significant differences between the conditions [$F(3, 114) = 3.78$, $p = 0.012$]. More particular, it appeared that both the request (M= 3.90, SD=0.83, p=.043) and content (M= 3.86, SD=0.87, p=.024) condition perceived the PFB request significantly more useful to formulate feedback, compared to students in the combination condition (M= 3.21, SD=1.14). No other significant differences were found compared with the control condition (M= 3.45, SD=0.87). Moreover, when a PFB request is part of the PA process, students believe they would provide more specific feedback in their PFB messages (M= 3.82, SD=0.83). Additionally, it appeared that these messages would comprise more suggestions on how to improve future work (M= 3.63, SD=0.80), revealing significant differences between the conditions [$F(3, 112) = 2.85$, p= 0.040]. Especially students in the content condition (M= 3.83, SD=0.65, p=.043) demonstrated they would provide significantly more suggestive feedback, compared to students in the combination condition (M= 3.28, SD=0.88). In comparison with the control (M= 3.64, SD=0.82) or request condition (M= 3.77, SD=0.77), no other significant differences were found. Similarly, students perceived the content checklist as a helpful instructional intervention to formulate PFB about a peers' draft version (M= 4.05, SD=0.80). In general, students more or less agreed that the content checklist actually helps the assessor to increase the quality of the PFB (M= 3.26, SD=0.95).

Discussion

Peer feedback quality

Taking into account that students had previous experience with PA and they did not perceive offering PFB as a difficult task, the results revealed that the PFB quality scores, measured by the Feedback Quality Index (Prins et al., 2006), increased for all students over time after multiple practice occasions. Also, being engaged in an online setting appeared not to be a problem for students, which supports research that acknowledges the beneficial effect of engaging students in online PA activities (eg. Cheng, Liang, & Tsai, 2015; Gielen & De Wever, 2015; Tsai & Chuang, 2013; Xiao & Lucking, 2008). However, only the control and the request condition increased significantly from time 1 to time 3. This finding suggests that by involving students in similar practice occasions, PFB quality scores could improve significantly over time. This is in line with earlier research (eg. Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010; Gielen & De Wever, 2015) claiming that practice is perceived as a condition sine qua non, in order to evolve as a skilled peer assessor. While there are studies indicating that students do not really require training in assessment (eg. Cho & MacArthur, 2010), most research stresses that feedback practices in higher education should take into account more practice occasions in providing and receiving of feedback (eg. Nicol, 2010). Although there were unexpected significant differences in PFB scores between the four conditions at the start of the intervention, results reveal that over time these significant differences are eliminated when students have multiple occasions to practice a similar performance.

As earlier research has shown that a certain amount of structure is beneficial for the PA process (eg. Gielen & De Wever, 2012; 2015), we take a closer look at what happens when students become more actively involved at time 2 in the PFB process, by focusing on the one hand on the assessee who formulates a specific peer feedback request for different criteria, and, on the other hand, on the assessor who prepares a content checklist as input source when formulating peer feedback. Completing a PFB request appeared to be a useful approach to increase the PFB quality, as both conditions that incorporated a PFB request in the PFB process had a significant increase in PFB quality scores from time 1 to time 2, with or without the assessor completing the content checklist. When the assessee has the opportunity to ask for specific PFB, the assessor could be more inclined to provide higher quality PFB on topic. More particular, findings showed that students in the content condition believed they would offer significantly more suggestive feedback, when a PFB request would have been implemented in their PFB process, compared to the students in the combination condition. One reason for this could be that students, who employed a content checklist, have collected a lot of information during their preparation. When responding to a particular PFB request, they believe they would provide more suggestive feedback in function of future improvement. On the other hand, we need to consider that students in the combination condition could perceive the PFB request as a restriction when composing their PFB messages. Related to this, students indicated not to feel

restrained when providing feedback on demand, and it became clear that they offered feedback on more aspects, than merely responding to the PFB request. This is in line with research of Gielen et al. (2010), which claims that the implementation of 'feedback on demand' could motivate and direct the assessor to provide more 'responsive' feedback (Webb, 1991), and thus provide PFB of a higher quality, driven by a higher 'individual accountability' and 'positive interdependence' (Slavin, 1981). From time 2 to time 3, only the control condition showed a significant increase while both conditions that incorporated a PFB request demonstrated a minor decline in PFB scores at time 3. This finding suggests that a PFB request could be useful approach to boost PFB quality scores, especially in the initial phase of performance, but that its effect could diminish over time. Related to this, it appeared that students generally acknowledged the usefulness of a PFB request during the assignment period. More particularly, the request and content condition perceived the PFB request significantly more useful compared to the combination condition. No significant differences were found in comparison with the control condition.

Future research should confirm whether the PFB request is actually able to boost the PFB scores, at least in the request condition, or whether this increase from time 1 to time 2 may be due to the significantly lower scores at time 1. Since students were randomly assigned to conditions, this different starting level was not expected and is a limitation of the study in the sense that it hinders a straightforward interpretation of the increase. Future research, replicating these conditions, may be necessary to confirm this hypothesis. When only the assessor is scripted to complete the content checklist before actually providing PFB, without being asked for specific feedback by the assessee, results pointed out no significant increase over time, taking into account that the content condition had generally high PFB quality scores at all times. Yet, results revealed no significant differences in PFB scores between the four conditions neither at time 2 or time 3. Also, given that both conditions that implemented the content checklist started at a rather high level (in spite of the random assignment of individuals to groups and the random assignment of groups to conditions), this may be the cause of the non-significant increases over time. Therefore, results are rather inconclusive to determine the actual effectiveness of the content checklist for the PFB process. Nevertheless, the questionnaire revealed that assessors perceive the content checklist as a helpful instructional tool, which they believe could have the potential to increase the assessor's PFB quality.

Product quality

In general, the questionnaire revealed that all students were generally satisfied with the quality of the PFB they received. Assessees indicated that their PFB was sufficient underpinned with specific article content or links to the article. As such, the received PFB was perceived as helpful to augment the quality of their writing performance. When looking into the product quality scores over time, there was an increase from time 1 to time 3 with respect to the writing products for all students. Similar to the PFB scores, this finding suggests that when students

have the opportunity to gain experience in a similar performance at multiple practice occasions, the quality of the product will increase generally as well. This result is in line with a previous study, which claims that practice leads to product improvement over time (Gielen & De Wever, 2015).

When decomposing the increase over time, results showed that the product scores of three out of four conditions increase significantly from time 1 to time 2, when assessee and assessor are more involved in the PFB process regarding a similar assignment. This is in agreement with research, which encourages the engagement of peers, both assessor and assessee in PA (e.g. Evans, 2013), as it requires students to deal with the essential aspects that correspond with high quality performance (Topping, 1998). Interestingly, completing a PFB request appeared to be a useful approach once again, not only to increase the PFB scores in the initial phase of performance, but as well to boost the product quality from time 1 to time 2, whilst both conditions that incorporated a PFB request had a significant increase in product scores, with or/and without the assessor completing the content checklist. This finding suggests that receiving specific feedback on request could have the potential to boost the quality of the peer feedback but also the quality of the actual writing product, since assessees could receive more detailed answers on particular previous issues. Gielen, et al. (2010) showed that the question form, in which secondary school students could request for feedback, led to more effective feedback. A possible explanation could be that "the assessee may in return pay more attention to the feedback that refers to these personal questions and this 'mindful reception' is crucial to the feedback's instructional value" (Gielen, et al., 2010, p. 158). This is in line with research, which suggests providing feedback on request as an approach to encourage students to make more effective use of their received feedback (Gibbs & Simpson's, 2004). When the assessor can give feedback on request, this could support the assessee to augment the quality of the (writing) product. This is in line with research of Shute, Hansen, and Almond (2007), showing that response - specific feedback appears to augment learning efficiency, while other research revealed that more specific and elaborated feedback stimulates better performance and outcomes (Strijbos, et al., 2010).

Since the product scores of the control condition showed a significant increase from time 1 to time 2 as well, without any additional structure or support, besides the provided structured peer feedback template, this finding may question the actual necessity of providing additional structure in the PFB process. This finding is in agreement with research, which advocates that every variety of feedback, whatever its amount or specificity, can have a positive effect on students' product scores (Topping, 1998). We should also take into account that all students, including the control condition, were using the same structured PFB template. In this way, the PFB process was already structured to a certain extent. From time 2 to time 3, no significant increase was found for any of the four conditions after one more practice occasion. Other research showed that PFB does not necessarily augments performance in a later phase of PA activity (Chen & Tsai, 2009), when students are involved in multiple practice occasions. Only the combination condition demonstrated a minor decline in PFB quality, but we have to take into account that their results were the highest of all conditions at time 2. It may be that after similar

practice occasions, the impact of the feedback request is diminishing. It appears that this strategy is especially useful to boost PFB and product scores in the initial phase of practice, but also that the effect diminishes after more practice occasions. This is in line with research, which claims that students' engagement in producing feedback reduces their own need for external feedback (Nicol, Thomson, & Breslin, 2014). For this reason, we believe that adaptable external collaboration scripts could be a possible solution, in which provided structure could be faded in or out over time, according to students' needs (eg. Wecker & Fischer, 2011). Following, we need to mention that the content condition had no significant increase from time 1 to time 2, or from time 2 to time 3, but this condition showed nevertheless a significant improvement from time 1 to time 3, and had even more the highest product scores of all conditions at the end of performance. Finally, results revealed no significant differences in product scores between the four conditions neither at time 2 or time 3.

Conclusion, limitations, and direction for future research

Taking into account, students perceived being engaged in online PA activities not as a difficult task, this study illustrated that offering additional structure in PA, to further specify the role of the assessee and assessor during the PFB process, is a valuable approach to increase both the quality of PFB and the writing product. Students indicated they were generally satisfied with the quality of the peer feedback, which appeared to be specific and elaborated with relevant information to improve future performance. Over time, results pointed out that all students improved after multiple practice occasions in providing PFB of a higher quality and delivering performances of a significant higher quality. It became clear that the additional structure for both assessee and assessor could be useful, especially in the initial phase of performance, but also that its effect diminishes over time. This study showed that students in their role as assessor especially appreciated the practicality of the PFB request and content checklist in the PA process. Overall, students were convinced that their received PFB was beneficial for the quality of their writing performance. For this reason, the data suggests that providing additional structure, such as a content checklist for the assessor and a PFB request for the assessee is a useful approach to actively engage students in writing and assessment activities and boost the quality of both PFB and product at the initial stage of performance.

Situated in an authentic learning environment, we have to take into account that not all contextual factors can be controlled for and that for example students' metacognitive approach, reading proficiency or computer literacy can affect the results. Although we tried to control this as much as possible by dividing students randomly over the different conditions, future research may consider taking into account these factors. As writing an abstract is a specific task, we have to consider that a significant improvement in product quality from draft to final version could be caused by the received PFB, but as well by experience through multiple practice occasions

(Kluger & Denisi, 1996). Related to this, detailed analyses of the specific content of the PFB request and checklist could shed more light on the PFB quality. Based on the recommendations of previous research (Gielen & De Wever, 2015), all students received an identical PFB template with a particular structuring degree for completing the third PFB phase. Thus, all conditions were structured to a certain extent in the PFB process. On one hand, structuring the PFB to a certain extent seems to be promising approach, given the overall increase in product scores over time, but, on the other hand, we are unable to fully determine which part of the product improvement is triggered by the quality of the peer feedback, by the received structure in the peer feedback process, or merely by practice in similar performance.

More research on optimizing the peer feedback process in function of students' learning is necessary. Future studies could replicate this study in order to examine the differential effect when the provided additional structure is kept or faded out at time 3, or when students have simply the option to create their own personalised script, in which they can select the required steps in the PFB process. As an implication for practice, instructors wishing to implement peer assessment could consider the following recommendations. Firstly, this study recommends implementing a structured PFB template (Gielen & De Wever, 2015) to guide students through the PFB process, in which they need to provide feedback and feed forward on a list of pre-specified, or preferably mutual discussed criteria (Sluijsmans, 2002), but they also need to evaluate the feedback in order to close the feedback loop (Boud, 2000), as illustrated in the appendix. Secondly, this study advises to foresee multiple practice occasions or feedback cycles, when instructors are planning to involve students as assessors and assessees in the PFB process.

References

Baker, M. J., & Lund, K. (1996). Flexibly structuring the interaction in a CSCL environment. In P. Brna, A. Paiva & J. A. Self (Eds.), *European conference on artificial intelligence in education* (pp. 401–407). Lisbon, Portugal.

Bangert-Drowns, R., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 7–74.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education, 38*, 698-712.

Butler, D. L. (2002). Individualizing instruction in self-regulated learning. *Theory and Practice, 41*, 81-92.

Chen, Y. C., & Tsai, C. C. (2009). An educational research course facilitated by online peer assessment. *Innovations in Education and Teaching International, 46*, 105–117.

Chen, Y. C., Liang, J. C. & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *Internet and Higher Education, 25*, 78–84.

Cho, K., & MacArthur, C. (2010) Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Cho, K. & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline. *Computers and Education, 48*, 409–426.

Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers and Education, 52*, 141–146.

De Wever, B., Hämäläinen, R., Voet, M., & Gielen, M. (2015). A wiki task for first-year university students: The effect of scripting students' collaboration. *The Internet and Higher Education, 25*, 37-44.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14*, 201-206.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research in computer-supported collaborative learning: From design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3–19). Netherlands: Springer.

Dochy, F., Segers, M., Van Den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*, 533-568.

Ebner, M., Kickmeier-Rust, M. & Holzinger, A. (2008). Utilizing Wiki-Systems in higher education classes: a chance for universal access? *Springer Universal Access in the Information Society, 7*, 199-207.

Ertmer, P. A., Newby, T. J., Liu, W., Tomory, A., Yu, H. E., & Lee, Y. M. (2011). Students' confidence and perceived value for participating in cross-cultural wiki-based collaborations. *Educational Technology, Research & Development, 59*, 213–228.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83*, 70–120.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66.

Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585-594.

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*. doi: 10.1111/jcal.12096

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.

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge and Taylor & Francis Group.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152.

Johnson, D. W., & Johnson, R. T. (1994). *Together and alone: Cooperative, competitive, and individualistic learning* (4th ed.). Boston, MA: Allyn & Bacon.

Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*, 387-406.

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment and Evaluation in Higher Education, 34*, 105–114.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice, 41*, 33–39.

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

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*, 159–185.

Kollar, I., & Fischer, F. (2010). Peer assessment as collabo- rative learning: A cognitive perspective. *Learning and Instruction, 20*, 344–348.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22-36.

Mostert, M., & Snowball, J. D. (2013). Where angels fear to tread: Online peer-assessment in a large first-year class. *Assessment & Evaluation in Higher Education, 38*, 674–686.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*, 501–517.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven prin- ciples of good feedback practice. *Studies in Higher Education, 31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review per- spective. *Assessment & Evaluation in Higher Education, 39*, 102–122.

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers and Education, 61*, 59-76.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465-477.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Sadeghi, H., & Kardan, A. A. (2015). A novel justice-based linear model for optimal learner group formation in computer-supported collaborative learning environments. *Computers in Human Behavior, 48*, 436–447.

Sadler, D. R. (1998) Formative assessment: revisiting the territory, *Assessment in Education, 5*, 77–84.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). *An assessment for learning system called ACED: Designing for learning effectiveness and accessibility*. (RR-07-26). Princeton, NJ: Educational Testing Service.

Slavin, R. E. (1989). Research on cooperative learning: an international perspective. Scandinavian *Journal of Educational Research, 33*, 231-243.

Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice* (2nd ed.). Boston, MA: Allyn & Bacon.

Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. J. G. & Bastiaens, T. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*, 23–42.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

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

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265–269.

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

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*, 20–27.

Tsai, C. C. (2009). Internet-based peer assessment in high school settings. In L. T. W. Hin, & R. Subramaniam (Eds.), *Handbook of research on new media literacy at the K-12 level: Issues and challenges* (pp. 743–754). Hershey, PA: Information Science Reference.

Tsai, Y. C., & Chuang, M. T. (2013). Fostering revision of argumentative writing through structured peer assessment. *Perceptual & Motor Skills, 116*, 210–221.

Tsai, C. C., & Liang, J. C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science, 37*, 293–310.

Tsai, C. C., Lin, S. S. J., and Yuan, S. M. (2002). 'Developing science activities through a networked peer assessment system', *Computers & Education 38*, 241–252.

Tseng S. C., Tsai C. C. (2010). Taiwan college students' self-efficacy and motivation of learning in online peer assessment environments. *Internet and Higher Education, 13*, 164–169.

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506-519.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5–13.

Van Lehn, K.A., Chi, M. T., Baggett, W. , & Murray, R. C. (1995). *Progress report: Towards a theory of learning during tutorin*g. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburg.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Berg, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher- order peer feedback in EFL. *Learning and Instruction, 20*, 316–327.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: the role of peer monitoring during the fading of instructional scripts. *Learning and Instruction, 21*, 746-756.

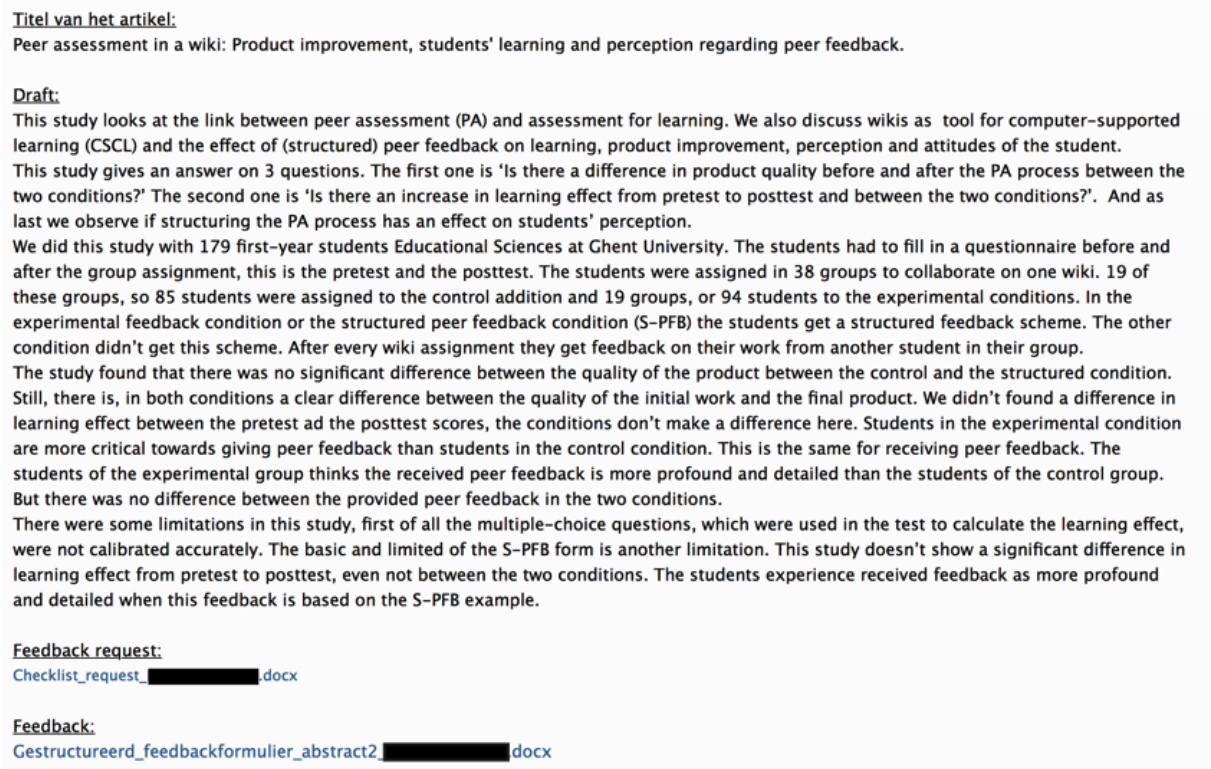
Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge—cognitive, computational and educational perspectives* (pp. 191–211). New York: Springer.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*, 186–193.

Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*, 72–83.

Appendix

Visual and chronological overview of all the different steps of one feedback cycle:



Titel van het artikel:
Peer assessment in a wiki: Product improvement, students' learning and perception regarding peer feedback.

Draft:
This study looks at the link between peer assessment (PA) and assessment for learning. We also discuss wikis as tool for computer-supported learning (CSCL) and the effect of (structured) peer feedback on learning, product improvement, perception and attitudes of the student.
This study gives an answer on 3 questions. The first one is 'Is there a difference in product quality before and after the PA process between the two conditions?' The second one is 'Is there an increase in learning effect from pretest to posttest and between the two conditions?'. And as last we observe if structuring the PA process has an effect on students' perception.
We did this study with 179 first-year students Educational Sciences at Ghent University. The students had to fill in a questionnaire before and after the group assignment, this is the pretest and the posttest. The students were assigned in 38 groups to collaborate on one wiki. 19 of these groups, so 85 students were assigned to the control addition and 19 groups, or 94 students to the experimental conditions. In the experimental feedback condition or the structured peer feedback condition (S-PFB) the students get a structured feedback scheme. The other condition didn't get this scheme. After every wiki assignment they get feedback on their work from another student in their group.
The study found that there was no significant difference between the quality of the product between the control and the structured condition. Still, there is, in both conditions a clear difference between the quality of the initial work and the final product. We didn't found a difference in learning effect between the pretest ad the posttest scores, the conditions don't make a difference here. Students in the experimental condition are more critical towards giving peer feedback than students in the control condition. This is the same for receiving peer feedback. The students of the experimental group thinks the received peer feedback is more profound and detailed than the students of the control group. But there was no difference between the provided peer feedback in the two conditions.
There were some limitations in this study, first of all the multiple-choice questions, which were used in the test to calculate the learning effect, were not calibrated accurately. The basic and limited of the S-PFB form is another limitation. This study doesn't show a significant difference in learning effect from pretest to posttest, even not between the two conditions. The students experience received feedback as more profound and detailed when this feedback is based on the S-PFB example.

Feedback request:
Checklist_request_[redacted].docx

Feedback:
Gestructureerd_feedbackformulier_abstract2_[redacted]docx

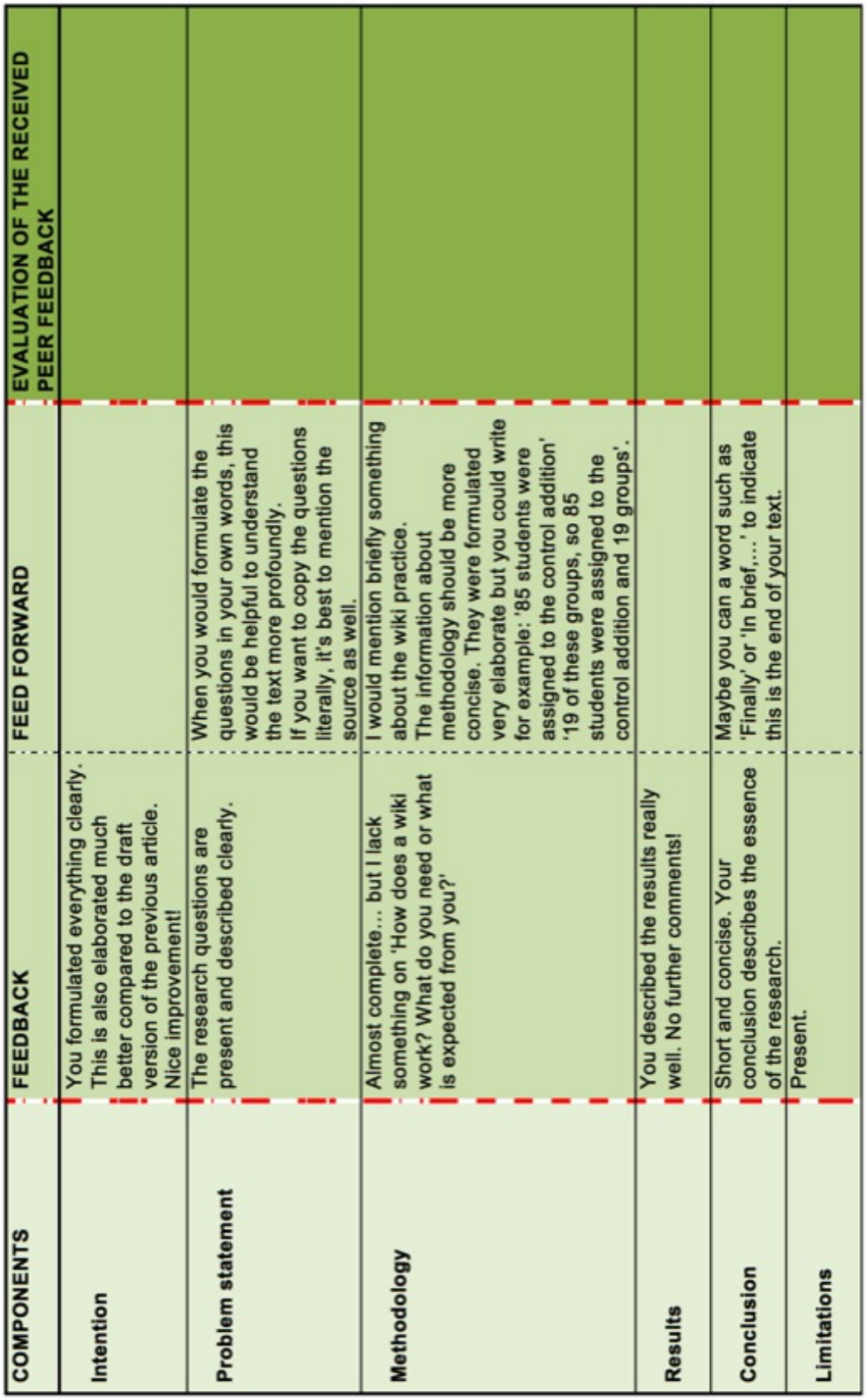
Phase 1: Draft version (The role of the assessee)

PEER FEEDBACK REQUEST	
CRITERIA	**REMARKS / QUESTIONS / ISSUES?**
INTENTION	• Should I also include something about the specific focus of this study? • ... • ... • ...
PROBLEM STATEMENT	• ... • ... • ... • ...
METHODOLOGY	• I wonder if I explain this section profoundly enough. How should I make it more concise because I believe it is rather long at the moment • Is this part structured enough? • ... •
RESULTS	• Did I overlook any important findings? Extra things to include? • ...
CONCLUSION	• Does my conclusion cover the message of the whole research study? • ... • ...
LIMITATIONS	• I cannot really find any limitations • Do you have a suggestion? • ... • ...
GENERAL	• Did I explain the specific terms well enough according to you? • ... • ... • ...

Phase 2: PFB Request (The role of the assessee)

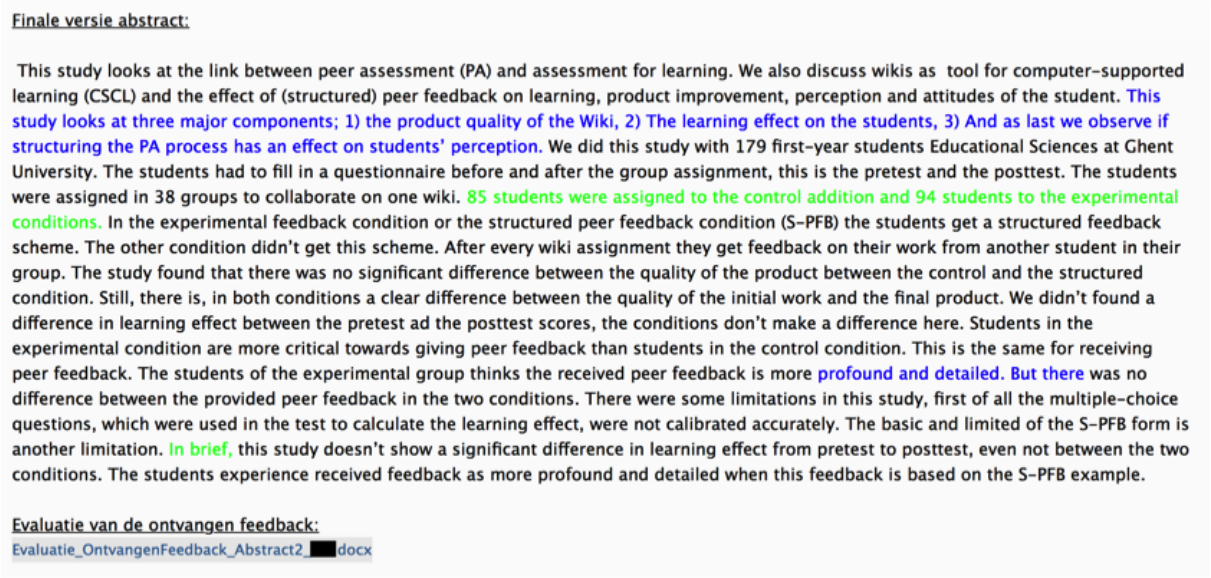
CONTENT CHECKLIST	
CRITERIA	**RELEVANT CONTENT FOR EACH CRITERIA**
INTENTION	• "...the goal of this paper is not so much to promote the use of commercial video games in education in se, as to understand, explain and predict changes in teachers' behavior in view of adopting these tools." • "The study contributes to an established body of research that has examined general reasons for playing video games ,the play behavior of teachers and teachers-in-training and teachers' acceptance of educational computer games." • "In this paper, a model-based approach to teachers' beliefs is presented and evaluated, based on the understanding that "teachers are faced with many variables that interact with each other to either facilitate or discourage the acceptance of technology"
PROBLEM STATEMENT	• "The present study focuses on the factors that influence the acceptance of commercial video games as learning tools in the classroom." • "When discussing teachers in relation to digital game-based learning, the focus is often on what they perceive as potential barriers to the implementation of games in their own practice." • "...to measure the concerns of the teachers regarding the difficulty of using games in their practice."
METHODOLOGY	• "...the focus is on teachers-in-practice." • "Secondary schools were contacted based on their denomination (i.e. community/subsidized public schools, and subsidized private schools), type of education (general, technical, and vocational) and geographical distribution." • "The teachers could fill in the questionnaires using the medium of their choice. This way, 505 teachers could be involved." • "The questionnaire consisted of three parts, examining demographic information, teacher related variables and the constructs of the research model." • "Demographic information included variables such as age and gender (0 female – 1 male). The teacher related variables included teaching experience (years), grade, and subject."
RESULTS	• "Most research variables are positively related. A notable exception is complexity, which relates negatively to personal innovativeness, behavioral intention, critical mass, experience and learning opportunities. While this negative relation was expected, considering the literature, the interrelations appear to be rather weak." • "High interrelations were found among behavioral

Phase 3: Content Checklist (The role of the assessor)



COMPONENTS	FEEDBACK	FEED FORWARD	EVALUATION OF THE RECEIVED PEER FEEDBACK
Intention	You formulated everything clearly. This is also elaborated much better compared to the draft version of the previous article. Nice improvement!		
Problem statement	The research questions are present and described clearly.	When you would formulate the questions in your own words, this would be helpful to understand the text more profoundly. If you want to copy the questions literally, it's best to mention the source as well.	
Methodology	Almost complete... but I lack something on 'How does a wiki work? What do you need or what is expected from you?'	I would mention briefly something about the wiki practice. The information about methodology should be more concise. They were formulated very elaborate but you could write for example: '85 students were assigned to the control addition' '19 of these groups, so 85 students were assigned to the control addition and 19 groups'.	
Results	You described the results really well. No further comments!		
Conclusion	Short and concise. Your conclusion describes the essence of the research.	Maybe you can a word such as 'Finally' or 'In brief,... ' to indicate this is the end of your text.	
Limitations	Present.		

Phase 4: Formulating PFB (The role of the assessor)



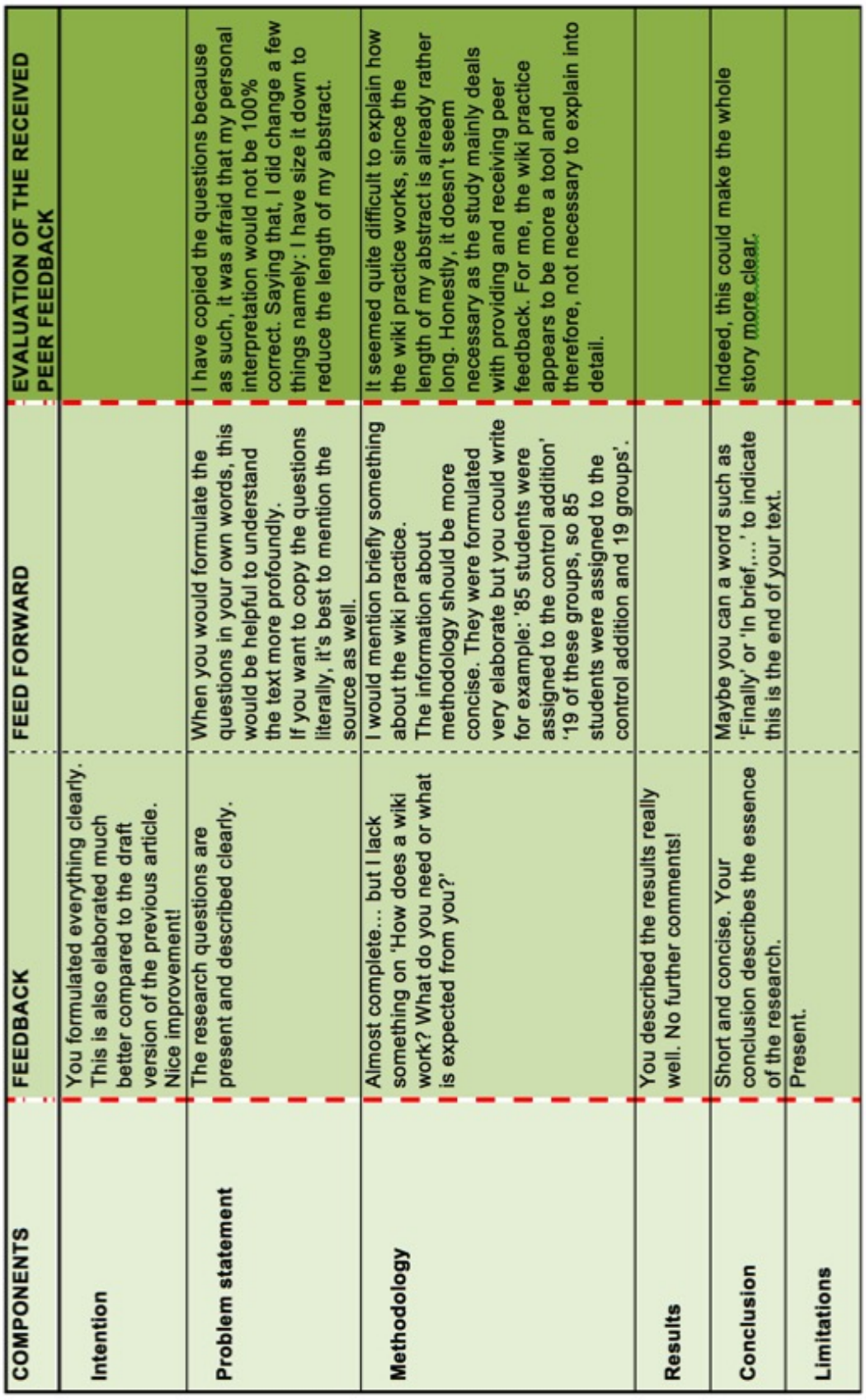
Finale versie abstract:

This study looks at the link between peer assessment (PA) and assessment for learning. We also discuss wikis as tool for computer-supported learning (CSCL) and the effect of (structured) peer feedback on learning, product improvement, perception and attitudes of the student. This study looks at three major components; 1) the product quality of the Wiki, 2) The learning effect on the students, 3) And as last we observe if structuring the PA process has an effect on students' perception. We did this study with 179 first-year students Educational Sciences at Ghent University. The students had to fill in a questionnaire before and after the group assignment, this is the pretest and the posttest. The students were assigned in 38 groups to collaborate on one wiki. 85 students were assigned to the control addition and 94 students to the experimental conditions. In the experimental feedback condition or the structured peer feedback condition (S-PFB) the students get a structured feedback scheme. The other condition didn't get this scheme. After every wiki assignment they get feedback on their work from another student in their group. The study found that there was no significant difference between the quality of the product between the control and the structured condition. Still, there is, in both conditions a clear difference between the quality of the initial work and the final product. We didn't found a difference in learning effect between the pretest ad the posttest scores, the conditions don't make a difference here. Students in the experimental condition are more critical towards giving peer feedback than students in the control condition. This is the same for receiving peer feedback. The students of the experimental group thinks the received peer feedback is more profound and detailed. But there was no difference between the provided peer feedback in the two conditions. There were some limitations in this study, first of all the multiple-choice questions, which were used in the test to calculate the learning effect, were not calibrated accurately. The basic and limited of the S-PFB form is another limitation. In brief, this study doesn't show a significant difference in learning effect from pretest to posttest, even not between the two conditions. The students experience received feedback as more profound and detailed when this feedback is based on the S-PFB example.

Evaluatie van de ontvangen feedback:

Evaluatie_OntvangenFeedback_Abstract2_[redacted]docx

Phase 5: Final version (The role of the assessee)



COMPONENTS	FEEDBACK	FEED FORWARD	EVALUATION OF THE RECEIVED PEER FEEDBACK
Intention	You formulated everything clearly. This is also elaborated much better compared to the draft version of the previous article. Nice improvement!		
Problem statement	The research questions are present and described clearly.	When you would formulate the questions in your own words, this would be helpful to understand the text more profoundly. If you want to copy the questions literally, it's best to mention the source as well.	I have copied the questions because as such, it was afraid that my personal interpretation would not be 100% correct. Saying that, I did change a few things namely: I have size it down to reduce the length of my abstract.
Methodology	Almost complete... but I lack something on 'How does a wiki work? What do you need or what is expected from you?'	I would mention briefly something about the wiki practice. The information about methodology should be more concise. They were formulated very elaborate but you could write for example: '85 students were assigned to the control addition' '19 of these groups, so 85 students were assigned to the control addition and 19 groups'.	It seemed quite difficult to explain how the wiki practice works, since the length of my abstract is already rather long. Honestly, it doesn't seem necessary as the study mainly deals with providing and receiving peer feedback. For me, the wiki practice appears to be more a tool and therefore, not necessary to explain into detail.
Results	You described the results really well. No further comments!		
Conclusion	Short and concise. Your conclusion describes the essence of the research.	Maybe you can a word such as 'Finally' or 'In brief,...' to indicate this is the end of your text.	Indeed, this could make the whole story more clear.
Limitations	Present.		

Phase 6: Evaluating PFB (The role of the assessee)

Tables

Table 1 - Multilevel models for the quality of the feedback (dependent variable: peer feedback score)

	Model 0	Model 1	Model 2	Model 3
Fixed				
Intercept (cons)	63.929(2.337)***	59.676(2.501)***	57.883(4.270)***	59.095(4.517)***
Time 2		5.952(1.540)***	5.952(1.540)***	1.594(2.980)
Time 3		6.808(1.540)***	6.808(1.540)***	7.531(2.980)*
Request - No Content			-5.526(5.883)	-10.100(6.359)
No Request - Content			6.519(6.084)	7.438(6.568)
Request - Content			6.952(5.941)	6.016(6.436)
Time2 * Request – No Content				10.285(4.183) *
Time3 * Request – No Content				3.438(4.183)
Time2 * No Request - Content				0.940(4.285)
Time3 * No Request – Content				-3.698(4.285)
Time2 * Request - Content				5.906(4.285)
Time3 * Request - Content				-3.098(4.285)
Random				
Level 3 - Group	103.230(40.341)*	103.230(40.341)*	77.100(33.431)*	77.100(33.431)*
ρ(%)	24.98%	25.55%	20.42%	20.64%
Level 2 - Student	147.978(29.301)***	152.562(29.236)***	152.295(29.168)***	154.343(29.141)***
ρ(%)	35.81%	37.76%	40.32%	41.31%
Level 1 - Time	162.029(14.492)***	148.278(13.262)***	148.278(13.262)***	142.133(12.713)***
ρ(%)	39.21%	36.69%	39.26%	38.05%
Model fit				
Deviance (-2LL)	3169.640	3147.468	3142.059	3131.477
$\chi 2$		22.172	5.409	10.582
df		2	3	6
p		*p*=.001***	*p*=.144	*p*=.102
Reference model		Model 0	Model 1	Model 2

Notes: *Significant at .05 level **Significant at .01 level ***Significant at .001 level Values in parentheses are standard errors

Table 2 - Multilevel models for the quality of the product (dependent variable: product score)

	Model 0	Model 1	Model 2	Model 3
Fixed				
Intercept (cons)	55.828(0.882)***	50.118(1.233)***	49.509(1.840)***	48.000(2.346)***
Time 2		7.504(1.493)***	7.504(1.493)***	9.063(2.929)**
Time 3		9.624(1.493)***	9.624(1.493)***	12.594(2.929)***
Request - No Content			-1.916(2.282)	0.333(3.293)
No Request - Content			2.192(2.337)	4.167(3.373)
Request - Content			2.492(2.337)	4.333(3.373)
Time2 * Request – No Content				-2.244(4.111)
Time3 * Request – No Content				-4.503(4.111)
Time2 * No Request - Content				-3.663(4.211)
Time3 * No Request – Content				-2.260(4.211)
Time2 * Request - Content				-0.363(4.211)
Time3 * Request - Content				-5.160(4.211)
Random				
Level 3 - Group	2.541(6.242)	2.541(6.242)	0.000(0.000)	0.000(0.000)
ρ(%)	1.28%	1.41%	0%	0%
Level 2 - Student	30.357(13.113)*	38.881(12.847)**	38.192(11.478)***	38.839(11.457)***
ρ(%)	15.36%	21.52%	21.53%	22.05%
Level 1 - Time	164.800(14.740)***	139.229(12.453)***	139.229(12.453)***	137.287(12.279)***
ρ(%)	83.36%	77.07%	78.47%	77.95%
Model fit				
Deviance (-2LL)	3036.956	2994.803	2990.305	2986.792
$\chi 2$		42.153	4.498	3.513
df		2	3	6
p		*p*=.001***	p=.212	p=.742
Reference model		Model 0	Model 1	Model 2

6 Structuring the role of both assessor and assessee during multiple feedback cycles: An attempt to enrich the content of students' peer feedback in a wiki environment

This chapter is based on:

Gielen, M., De Wever, B. & Voet, M. (Accepted). Structuring the role of both assessor and assessee during multiple feedback cycles: An attempt to enrich the content of students' peer feedback in a wiki environment. *Journal of Computer Assisted Learning.*

Chapter 6
Structuring the role of both assessor and assessee during multiple feedback cycles: An attempt to enrich the content of students' peer feedback in a wiki environment

Abstract

This study investigates whether an instructional intervention in which both the assessor and assessee are actively involved in the peer feedback process enhances students' peer feedback content, within the context of writing assignments in a wiki-based computer-supported collaborative learning environment in the first year of higher education. The main aim was to find out if further structuring the role of assessee and assessor, on top of providing peer feedback with the help of a structured peer feedback template, has a beneficial impact on the peer feedback content. In this particular study, the assessee is required to compile a specific peer feedback request, while the assessor is asked to complete a content checklist as a preparation in the peer feedback process. The effect of this increased support was investigated using a 2x2 factorial design. Content analysis was conducted to compare students' peer feedback messages in four conditions: a (1) control, (2) request, (3) content, and (4) combination (i.e. request + content) condition by focusing in more detail on the following peer feedback content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, (5) elaboration focus, (6) evaluation agreement, and finally (7) evaluation implementation. Results revealed that all conditions provide a rather balanced proportion of mostly positive verifications and equally informative and suggestive elaborations. When evaluating their received peer feedback, students mainly tend to agree with the comments (62% of the cases), and actually implement the feedback in 41% of the cases. Although no significant differences between the conditions were found, this study stresses why instructors still could actively involve all actors in the PFB process, with the underlying purpose to increase the potential impact of PA and boost students' learning in higher education. Building on this, limitations, practical implications, and directions for future research are presented.

Introduction

Promoting students' learning and performance through online peer assessment

A large body of research has suggested that assessment can be valuable for students' learning (e.g. Carless, 2015; Garcia, Garcia-Alvarez & Moreno, 2014; Sambell, McDowell, & Montgomery, 2013), especially in constructivist-oriented learning environments where learners seek, give, obtain and act on feedback (Price, Handley, Millar, & O'Donovan, 2010), instead of merely receiving an external evaluation (Butler & Winne, 1995). Such practices of assessment *for* learning draw heavily on formative assessment procedures, such as peer assessment (PA), in order to diminish the discrepancy between learners' current and desired performance (Sadler, 1989). During PA, learners are expected to evaluate the quality of their peers' performance (Topping, 2009). When peer assessment is complemented with peer feedback, students are generally expected to provide feedback and revise their own work based on feedback messages from peers (Topping, 2009). When it comes to formulating feedback, students have to express feedback on how well the performance of same-level peers corresponds with the expected outcome, but also to express opinions, ideas and suggestions on how future performance can be enhanced (Hattie & Timperley, 2007).

Various studies have indicated that peer feedback can have a positive impact on different aspects of students' learning process and performance. In this respect, a growing body of research illustrated that involving learners in online peer feedback activities appears to augment the quality of their writing performance (e.g. Gielen & De Wever, 2015a; 2015c; Tsai & Chuang, 2013; Xiao & Lucking, 2008). However, studies have also demonstrated that the effectiveness of feedback is largely determined by its content, form and function (Narciss, 2008). In this light, previous work stresses that not all learners are able to give good feedback, and that this might be due to limited competence, lack of objectivity, or insufficient knowledge of how to interact with others (Cheng, Liang, & Tsai, 2015; Strijbos, Narciss, & Dünnebier, 2010). Therefore, further research on determining the required support during the peer feedback process is essential for optimizing the feedback that learners provide to one another (Hovardas, Tsivitanidou, & Zacharia, 2014).

Increasing engagement in PA through structuring the roles of assessor and assessee

Previous research revealed that most learners need instructional support to engage in high-level cognitive processes, such as critical thinking and problem solving (King, 2002), which appeared to be an essential aspect of actively participating in PA (Cole, 2009). One approach that

has been found to be effective within the context of wiki-based tasks (e.g. Gielen & De Wever, 2015c), consists of structuring students' work with collaboration scripts. Grounded in a scripted cooperation approach (O'Donnell, 1999), collaboration scripts are instructional supports that structure students' collaboration process, by specifying the roles and activities of all of the actors involved (Fischer, Kollar, Stegmann, & Wecker, 2013). Previous research suggests that structuring students' work is effective for optimizing the PA process and might even be a necessary precondition for effective feedback (Poverjuc, Brook, & Wray, 2012). For this reason, this study examines the effect of structuring the PA process on the actual PFB content. In a previous study, a PFB template with a varying structuring degree was tested out to examine the effect on the actual PFB content. Following this, the present study uses an optimised structured PFB template, based on the findings of the previous study (Gielen & De Wever, 2015a), and focuses on the question of how the PFB content evolves when both the assessor and assessee are increasingly involved in PFB process. In this light, it is interesting to examine if the outcome of the PFB process improves even further when all actors receive a higher degree of structuring.

With respect to this question, the literature provides various suggestions on how to structure the role of the assessor can be found in the literature (for more information, see also Van Merriënboer, Kirschner, & Kester, 2003). Until now, a growing body of research has experimented with varying instructional interventions in the PA process to enhance the effectiveness of PFB, for instance, by: organizing a training to improve PFB (Sluijsmans et al., 2002), working with multiple raters instead of one (Cho & Schunn, 2007), offering guiding questions to support the assessor while formulating PFB (Gielen & De Wever, 2012), providing sentence openers to encourage interaction between students (Baker & Lund, 1996), or by creating a PFB template with a varying structuring degree to provide feedback and feed forward (Gielen & De Wever, 2015a). Interestingly, almost all these instructional interventions mainly concentrate on the role of the assessor, while the role of the assessee is often forgotten in the instructional collaboration scenarios for peer assessment and feedback (Gielen, et al., 2010). However, previous research stressed that a collaboration script organizes the collaboration process of all involved actors, which means that it would be worthwhile to pay attention to the role of the assessee as well (Fischer, Kollar, Segmann, & Wecker, 2013). Therefore, we are interested in *what impact additional structure in the PFB process – not only for the assessor, but for the assessee as well - can have on the actual PFB content*.

Structuring the role of the assessee: Asking and evaluating PFB

In an attempt to include all actors more actively, previous research often proposes to encourage the assessee to formulate on which particular components feedback is required (Nicol & MacFarlane-Dick, 2006; Gibbs & Simpson, 2004). In this way, the assessee can be involved in the initial phase of the PFB process by compiling a '*feedback request*' (Gielen & De Wever, 2015c), also referred to in the literature as '*a priori question form*' or '*feedback on demand*' (Gielen et al., 2010). By incorporating a PFB request, assessees have the opportunity to formulate questions on a few particular criteria, ask advice for specific queries, or request

confirmation on uncertainties, while the assessors are in turn encouraged to generate more directive feedback. With regards to the reception of feedback, previous research stressed that it is important to take into account the response of the assessee in a final phase of the feedback cycle (Gibbs & Simpson, 2004), in order to close the feedback loop (Boud, 2000). Previously, other research referred to '*a posteriori reply form*' (Gielen et al., 2010) or 'back-feedback' (Kim, 2009). In this view, it is essential that the assessee evaluates the received PFB, in order to give an indication if the assessor's feedback was relevant, helpful, implementable, etc. In this way, assessees feel actively involved during the entire PFB process, since they have the opportunity to reflect upon the received PFB, which generates a so-called feedback dialogue (Prins & Mainhard, 2009). Figure 1 illustrates the different steps of how both the assessee and assessor can be activated in the PFB cycle.

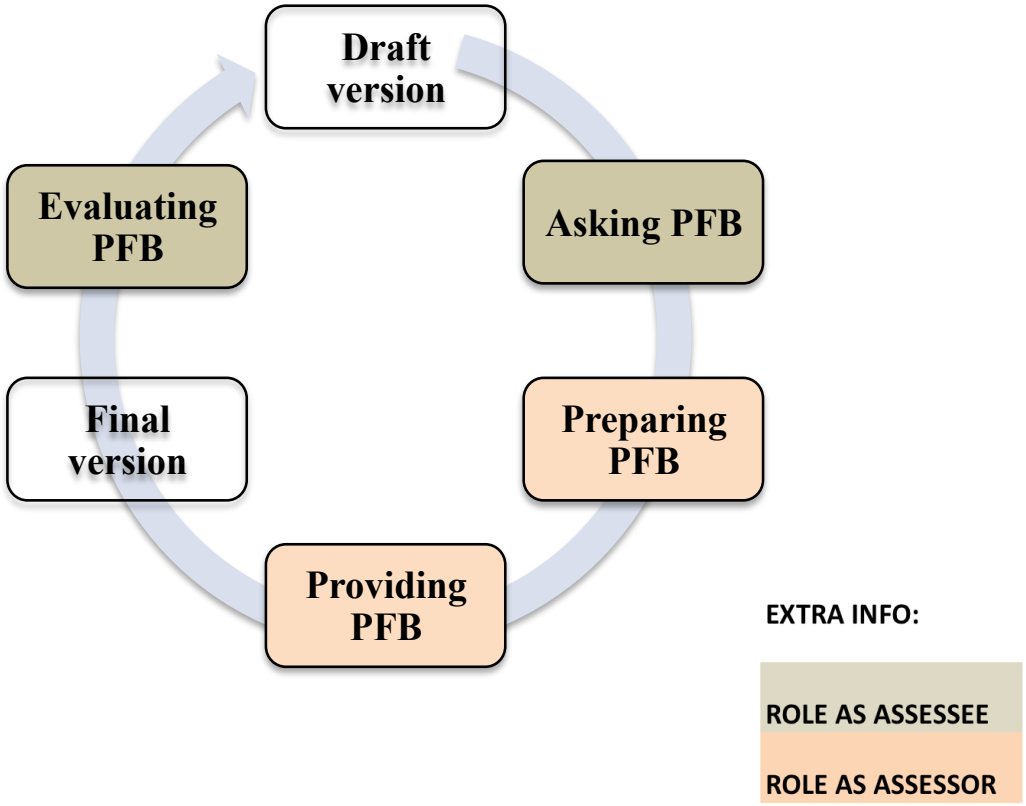


Figure 1. Overview of the Peer Feedback Cycle

Structuring the role of the assessor: Preparing and providing PFB

In order to encourage assessors to think carefully about how the expected performance of the assessee should look like or which elements should be present, guidelines and checklists appeared to be beneficial for the PFB process (Topping, 2009). As such, assessors could prepare themselves before actually seeing the product that the assessee has developed, leading to more different ideas, and avoiding the risk of assessors thinking in the same patterns of the developer of the project, and thus following his footsteps without considering different options and

identifying shortcomings. Assessors will be steered to profoundly process a peers' work and prepare the feedback accordingly (Kollar & Fischer, 2010), before actually formulating the feedback. Like this, a checklist can be used as starting point and input source for actually providing feedback.

Analysing peer feedback content

Essentially, feedback should answer three questions: 'Where am I going?' (feed up), 'How am I going?' (feed back), and finally 'Where to next?' (feed forward) (Hattie & Timperley, 2007). Traditionally, the feedback framework distinguishes between verifications and elaborations (Narciss, 2008). While verifications merely indicate how well particular criteria were achieved or not, elaborations provide more details on what was the reason why something was right or wrong and secondly, what students can do to enhance the quality of their performance (Hattie & Gan, 2011). Previous work suggests that all students habitually provide more positive verifications in their PFB messages, but also that an instructional PFB template with a higher structuring degree can result in more negative comments (Gielen & De Wever, 2015b). Stressing the importance of elaboration, research emphasizes that feedback should provide the assessee with improvement strategies (Duijnhouwer, Prins, & Stokking, 2012). This is central to formative assessment (Sadler, 1989), and contributes to students' writing performance (e.g. Stern & Solomon, 2006). When the assessor provides additional informational feedback in support of a specific evaluation, studies claim that justifications are most beneficial for students' performance (eg. Walker, 2015; Webb & Mastergeorge, 2003). However, other research claims that feedback should include suggestions as well in function of future improvement (e.g. Butler, 1987).

As feedback should address particular aspects of the task (Kluger & DeNisi, 1996), research revealed that task-specific feedback appears to be beneficial for performance (Hattie & Timperley, 2007). More specifically, feedback can direct the attention of the assessee to certain mistakes, a particular topic or answer, but feedback can also provide more general support or guidance to assist the assessee (Shute, 2008). As 'mindful reception' of the feedback contributes to students' learning (Salomon & Globerson, 1987), recent research emphasised that it would be helpful to also look at how students actually deal with the feedback (Walker, 2015). Assessees can react differently towards external feedback. For instance, assessees may simply ignore and reject external feedback, because they perceive it as irrelevant, and not in line with internal feedback (Butler & Winne, 1995). For this reason, research should focus more on the role of the assessee who needs to process and employ the feedback (Molloy & Boud, 2014). In this view, this study examines not only if assessees generally agree with the received PFB messages, but also we attempt to gain more insight on whether the assessees habitually implement the feedback in their writing assignments.

Hypotheses

Based on the literature review above, we formulate the following hypotheses for this study:

H1 - While verifications merely state if something is right or wrong (Hattie & Gan, 2011), elaborations contain more relevant information to assist the assessee. Good feedback comprises both the verification and elaboration component (Mason & Bruning, 2001). We assume that all students will provide in their PFB messages a balanced proportion of verifications and elaborations (H1a), not revealing significant differences between the conditions (H1b).

H2 – Previous research revealed that both positive and negative feedback might increase and decrease motivation and performance (Van Dijk and Kluger, 2004). When learners receive positive feedback, this could result in "sitting on their laurels", but as well in "doubling their efforts". Similarly, when learners receive negative feedback, this could lead to "giving up", but as well to "trying harder". In line with earlier findings, we predict that all students will provide predominantly positive verifications in their PFB messages (H2a), in which these verifications largely focus on particular criteria (H2b), due to the help of a structured PFB template for the assessor, not revealing significant differences between the conditions (H2c), since all conditions will use the same template.

H3 – When students elaborate on previously given verifications, we hypothesize that all students will provide a balanced proportion of informative and suggestive elaborations (H3a), that are general elaborations that are focused on particular criteria (H3b) as feedback should address particular aspects of the task (Kluger & DeNisi, 1996). Previous research indicated that students provide generally more suggesting than informative elaborations, independently from the structuring degree of the PFB template (Gielen & De Wever, 2015b). However, we hypothesize that students will provide more informative elaborations, when the assessee and/or assessor receive additional structure in the PFB process, (H3c).

H4 - In order for PFB to be effective, assessees are required to engage with the feedback (Molloy & Boud, 2014). We hypothesize that students will generally agree with most of the received feedback (H4a), and consequently implement most of the received feedback (H4b). When assessees receive well-prepared PFB with the help of the 'content checklist' and/or personalised PFB through 'feedback-on-demand', we hypothesize that students who receive additional structure in the PFB process, will agree more with the received feedback (H4c), and therefore, would be more inclined to apply the feedback (H4d) (Gielen, et al., 2010).

Material and methods

Participants, research design and conditions

In total, 125 first-year university students, enrolled in an educational sciences program, were divided into 27 groups of maximum five students. In every group, each member was instructed to write an abstract for three academic journal articles, from which the abstract was removed. Fifteen different submitted, but not yet published, research articles were selected, so that each group member had 3 different articles. As part of the curriculum requirements, each group member was structured to (1) write a draft version of the abstract, (2) formulate PFB on the draft version of a fellow student, (3) construct a final version of their own abstract, inspired by the received PFB, and finally (4) evaluate the received peer feedback. This process was repeated over three cycles. When providing PFB, all students employed an identical structured PFB template, which was based on the findings of Gielen & De Wever (2015a), listing seven criteria which the assessor needs to take into account when providing PFB (intention of the research, problem statement, methodology, results, conclusion, limitations, and general judgments). Additionally, the PFB template also structured the assessors to formulate not only feedback but also feed forward for each of the seven criteria (Hattie & Timperley, 2007).

INDEX	PROCEDURE	CONTROL	REQUEST	CONTENT	COMBINATION
	START CYCLUS X				
1	READING ARTICLE	✔	✔	✔	✔
2	WRITING DRAFT VERSION	✔	✔	✔	✔
3	PEER FEEDBACK REQUEST		✔		✔
4	PREPARING CONTENT CHECKLIST			✔	✔
5	PROVIDING PEER FEEDBACK	✔	✔	✔	✔
6	WRITING FINAL VERSION	✔	✔	✔	✔
7	EVALUATION PEER FEEDBACK	✔	✔	✔	✔
	END CYCLUS X				

EXTRA INFO:

ROLE AS ASSESSEE

ROLE AS ASSESSOR

Figure 2. Overview of the required procedure for each condition

At time 1, the only requirement all conditions had to fulfill during the PFB process, was to employ the structured PFB template. In function of the four conditions, additional structure was intentionally offered for the assessee and the assessor at time 2 and time 3. The instruments used during the PFB process, will be discussed into more detail in the next section. This study applied a quasi-experimental 2 x 2 factorial design, which resulted in four conditions. In the *feedback request* condition, the assessee had to complete a feedback request to clarify on which

particular aspect of the assignment they required peer feedback. In the *content checklist* condition, the assessors had to complete a checklist based on article content before formulating peer feedback on the draft. In order to do so, students were required to read the assessee's article, select and categorize the essential content within a list with separate bullet points for each criterium. Both types of support are joined in the *combination condition*, which requires the assessee to formulate a PFB request and the assessor to prepare the PFB through the help of a content checklist. In the *control condition*, students were also required to use the structured PFB template at all times, but assessees did not need to specify their peer feedback request, nor did assessors need to prepare a checklist before formulating peer feedback. Figure 1 provides an overview of the specific procedure for each condition.

Instruments

Peer feedback request

In the feedback request condition, the assessees were requested to formulate 'feedback on demand' (Gielen, et al., 2010) after submitting the draft version of an abstract. This required them to evaluate their own work, and to think about the particular aspects of the task for which they would like to receive PFB. Next to this, the assessees did not receive specific instructions on how to formulate a feedback request, but were instead left free to formulate them as they best saw fit, as illustrated in the appendix.

Content checklist

In the content checklist condition, assessors were required to complete a content checklist as preparation, before actually formulating PFB. More specifically, these students were asked to thoroughly read the article assigned to the assessee, and then select the information that should be included in the abstract. Following this, the assessor was required to categorise the selected article content according to the seven predetermined criteria, as illustrated in the appendix. Assessors were required to finalize the content checklist before reading the draft version of the assessee, as the checklist was to used as a basis for formulating peer feedback.

Structured PFB template

Based on specific suggestions of Gielen and De Wever (2015a), the same structured PFB template was offered to all four conditions during the assignment period. As illustrated in Figure 3, this structured PFB template consisted of four different sections. First of all, a list of seven predetermined criteria (section 1) made it clear on which aspects students should focus during the PFB process. Next, the assessors needed to provide feed back for each criterion (section 2 - How are you going?), by indicating how well the draft corresponded to expectations, together

with feed forward (section 3 – Where to next?), containing ideas and suggestions in order to increase the quality of the abstract. In a last step, assessees were required to evaluate their received PFB (section 4) to give an indication about the value and relevance of the received PFB, in order to close the feedback loop (Boud, 2000).

COMPONENTS	FEEDBACK	FEED FORWARD	EVALUATION OF THE RECEIVED PEER FEEDBACK
Intention	You formulated everything clearly. This is also elaborated much better compared to the draft version of the previous article. Nice improvement!		
Problem statement	The research questions are present and described clearly.	When you would formulate the questions in your own words, this would be helpful to understand the text more profoundly. If you want to copy the questions literally, it's best to mention the source as well.	I have copied the questions because as such, it was afraid that my personal interpretation would not be 100% correct. Saying that, I did change a few things namely: I have size it down to reduce the length of my abstract.
Methodology	Almost complete… but I lack something on 'How does a wiki work? What do you need or what is expected from you?'	I would mention briefly something about the wiki practice. The information about methodology should be more concise. They were formulated very elaborate but you could write for example: '85 students were assigned to the control addition' '19 of these groups, so 85 students were assigned to the control addition and 19 groups'.	It seemed quite difficult to explain how the wiki practice works, since the length of my abstract is already rather long. Honestly, it doesn't seem necessary as the study mainly deals with providing and receiving peer feedback. For me, the wiki practice appears to be more a tool and therefore, not necessary to explain into detail.
Results	You described the results really well. No further comments!		
Conclusion	Short and concise. Your conclusion describes the essence of the research.	Maybe you can a word such as 'Finally' or 'In brief,…' to indicate this is the end of your text.	Indeed, this could make the whole story more clear.
Limitations	Present.		

Figure 3. Structured Peer Feedback template

Data for content analysis and unit of analysis

In order to analyse the specific content of PFB messages, a random subsample of 16 groups (four groups from each condition) out of 27 groups in total, was selected. As all three PFB cycles were analysed, this resulted in 237 peer feedback forms from 79 students in total. Based on a content analysis scheme, which was also used in the preceding study, the specific content of PFB messages was split up into smaller fragments. Inspired by strategy of Strijbos, Martens, Prins, and Jochems (2006), a segmentation of the data preceded the actual coding process. The segmentation procedure of Strijbos et al. (2006) was used to divide the specific content of PFB messages into segments. Similar to Rourke, Anderson, Garrison, & Archer (2001), this study works with the syntactical unit or sentence level, as it is more feasible to categorise sentences or part of sentences into the coding scheme. After the segmentation and coding process, the 237 peer feedback forms resulted in a database of 8440 segments.

Coding scheme for content analysis

This study employs a recently developed coding scheme (Gielen & De Wever, 2015b) that was inspired by a coding scheme for analysing peer feedback messages (Strijbos, et al., 2012), which in turn builds on the feedback framework of Narciss (2008). This coding scheme attempts to identify variations in the quality of peer feedback content, by concentrating on the peer feedback style, type, and focus. In addition, it also takes into account students' evaluation of the received PFB, by looking at their agreement with and implementation of the feedback. Identical to Strijbos, Van Goozen, & Prins (2012), this content scheme incorporates the categories '*feedback style'* (verification – elaboration – general) and '*verification type*' (positive – negative – neutral). Regarding '*elaboration type*', we distinguish between informative elaborations, in which the feedback of peers informs, evaluates, confirms or justifies how well past performance was, and secondly suggestive elaborations, in which suggestions, ideas, advice, etc. are provided in function of future performance. Regarding the verification and elaboration focus, the coding scheme divides between general or specific feedback on the overall assignment, on particular criteria or on language features.

In order to close the feedback loop (Boud, 2000), the evaluation of the received PFB was taken into account to examine whether the assessee agrees with the received PFB and secondly, if the assessee is indicating that (s)he will implement the suggestions that are formulated in the PFB. For this particular study, the categories '*Evaluation Agreement' and 'Evaluation Implementation?'* were added to the originally developed content analysis scheme of Gielen & De Wever (2015b). Appendix 1 presents and provides examples of these seven coding categories. Both tests on segmentation and interrater reliability were satisfactory. As all Kappa values were above the benchmark of .80 (Landis & Koch, 1977), there was a high agreement for all categories (For more information, see Gielen & De Wever, 2015b).

Data analysis strategy

As there were three measurement occasions (i.e. the peer feedback moments, indicated by the variable 'time', level 1), which are nested within each of the 79 students (level 2), who are in turn nested within 16 groups (level 3), multilevel modelling (MLwiN 2.29) was used to tackle this hierarchical structure of the data. For all categories of the content analysis scheme, multilevel analyses were performed to compare the effect of the conditions with respect to participants' peer feedback content quality. Initially, we examined the null model for each of the categories of peer feedback content to confirm whether the multilevel approach was acceptable, compared to a single-level regression analysis. Afterwards, a compound symmetry model was estimated, by adding the categorical predictor 'time' to the null model (Snijders & Bosker, 1999). In this model, the two last measurement occasions (time 2 and 3) can be compared with the reference category (time 1). The categorical predictor 'condition' was added in the next step.

Subsequently, the interaction 'condition*time' was included in the model. In a final phase, the number of segments was added to the model, to take into account the (significant) difference(s) between the four conditions regarding the number of segments. Furthermore, logistic regression analyses confirmed that using percentages was justified during multilevel analyses.

Results and Discussion

As shown in Figure 4, the results show that the proportion of verifications and elaborations in feedback messages was rather balanced within all conditions had a rather balanced percentage of, with slightly more elaborations, confirming H1a. This is in line with a previous study, which found similar results (Gielen & De Wever, 2015b). As such, this finding suggests that students did not only indicate whether certain criteria were met, but also that they took care to explain these evaluations.

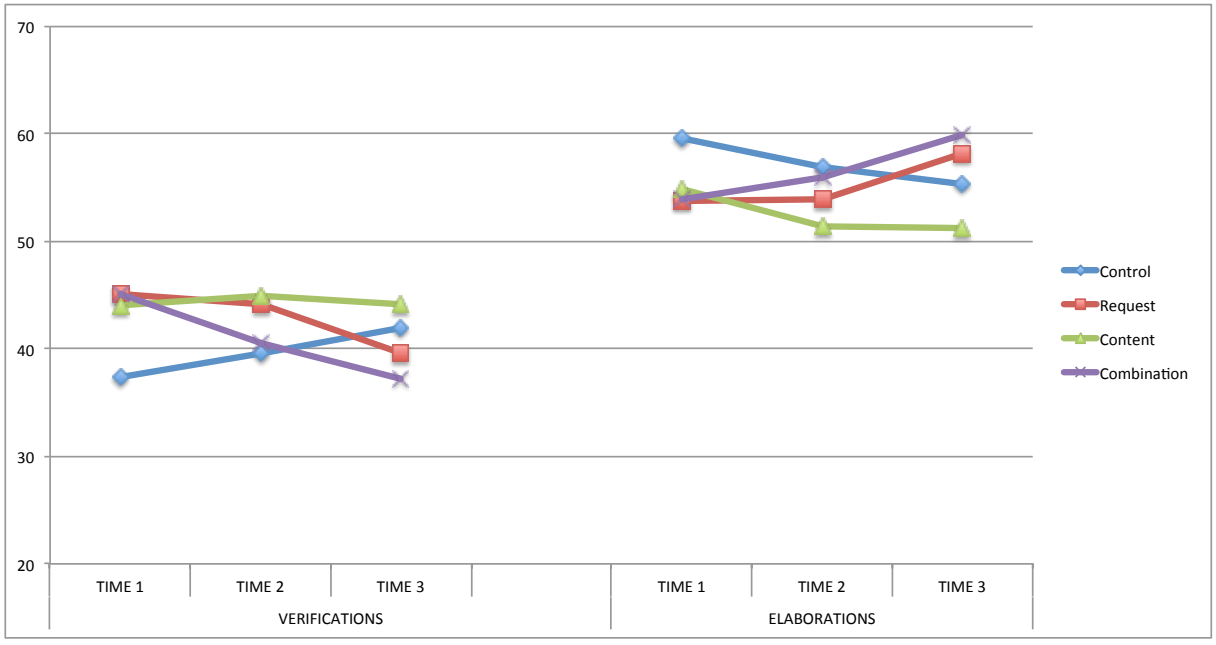


Figure 4. An overview of percentages of verifications and elaborations over time

Although the control condition started off with a significantly lower percentage of verifications in their PFB messages at time 1 (M=37%), compared to the request (M=45%, p=.010), content (M=44%, p=.025) and combination (M=45%, p=.010) condition, no significant

differences were found at time 2 or time 3. Over time, results revealed a significant decrease from time 1 to time 3 in verifications for the request (-5,4%, *p*=.030) and combination (-7,9%, *p*=.002) condition. Related to this, the elaboration percentage increased significantly for the combination condition (+5,8%, *p*=.029) from time 1 to time 3, which resulted in significantly more elaborations for the combination condition (M=60%), compared to the content condition (M=51%, *p*=.025). Next to this, no significant differences were found between the conditions, which only partly confirms H1b. This finding suggests that when a PFB request is incorporated over time in the PFB process, students tend to provide less verifications, and formulate hence a higher percentage of elaborations in their messages. When students provide 'feedback on demand' in order to tackle a particular error, topic or response (Shute, 2008), this finding suggests that students will verify less if something is merely right or wrong in their PFB messages, but formulate more informative and suggestive comments to justify their respond towards the assessee in order to meet this particular criteria.

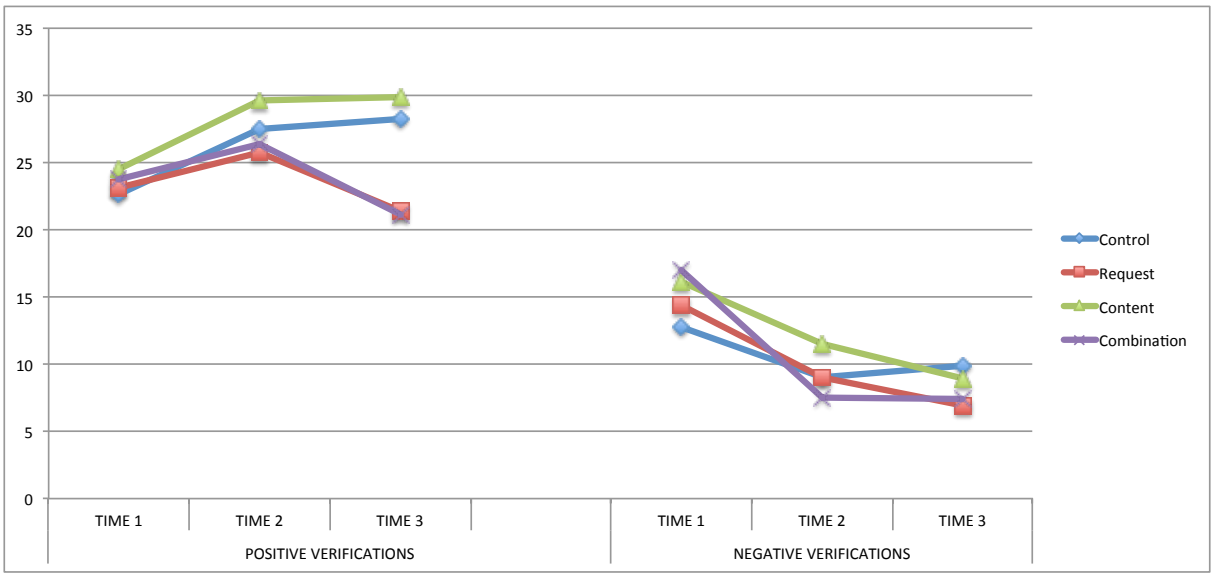


Figure 5. An overview of percentages of positive and negative verifications over time

Taking a closer look at the verification component, Figure 5 illustrates that all students provide more positive than negative comments. This finding is in line with the findings of the preceding study (Gielen & De Wever, 2015b) and confirms H2a. Due to the structured PFB template's focus on specific criteria, more than 80% of these positive or negative comments were verifications that are focused on particular criteria, confirming H2b. Due to triviality, the category 'verification focus' and its subcategories will not be reported, but only illustrated in the appendix. In order to find out more about the differential impact of additional structure, we take a closer look at differences between the conditions. Results showed that the control (+5,7%, *p*=.031) and content (+5,4%, *p*=.038) condition provide significantly more positive comments from time 1 to time 2. On the other hand, results indicated a significant decline in negative comments from time 1 to time 2 for the request (-5,2%, *p*=.031) and combination (-9,4%,

p<.001) condition. Finally, there was a significant decline of negative comments from time 1 to time 3 for the request (-7,4%, *p*<.001), content (-7,2%, p<.001) and combination (-9,6%, *p*<.001) condition. However, results indicated certain fluctuations over time, no real significant differences between the conditions were found, confirming H2c).

This finding suggests that students are inclined to formulate more positive than negative comments. Without ignoring the importance of positive feedback, students have to be confronted to formulate negative feedback as well, as this may result in increased effort (e.g. Bandura & Cervone, 1986) and may be necessary to reveal inadequacies in a peer's performance. Related to this, the preceding study illustrated that assessors provide more negative comments when they employ a PFB template with a higher structuring degree (Gielen & De Wever, 2015b). Over time, it appeared that additional structure for the assessee and assessor, complementary to the structured PFB template for all students, could generate significantly less negative comments in their PFB messages after multiple feedback cycles. We should however be aware that another reason could be that students' work could simply require less negative comments over time, as performance improves after multiple practice occasions in similar writing assignments.

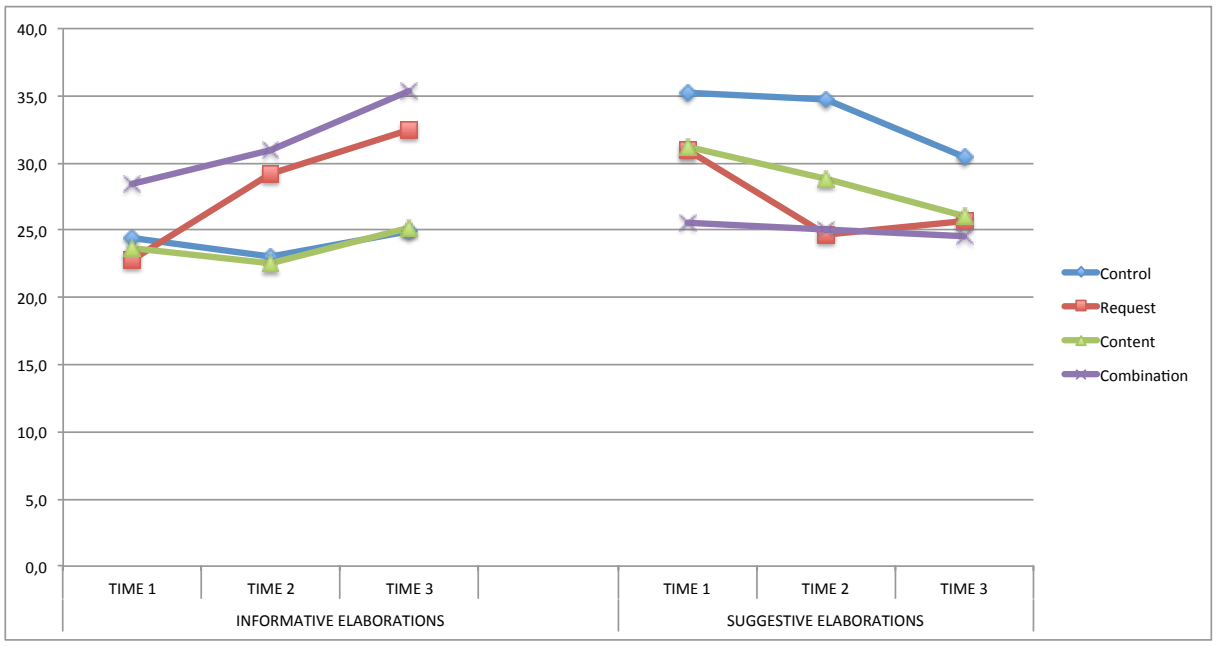


Figure 6. An overview of percentages of informative and suggestive elaborations over time

Regarding the elaboration component of the feedback, results in Figure 6 indicate that students provide generally a balanced proportion of informative and suggestive elaborations, confirming H3a, which is in line with the findings of the preceding study. This finding implies that all students justify in their PFB messages why certain criteria were met or not, but they also suggest how a peer's draft version can be improved, which appears to be beneficial for its effectiveness (e.g. Hattie & Gan, 2011; Butler, 1987). Similar to verification focus, results

indicated that more than 80% of these informative or suggestive comments are general elaborations that are focused on particular criteria, caused most probably by the structured PFB template, which confirms H3b. Due to triviality, the category 'elaboration focus' and its subcategories will not be reported, but only illustrated in the appendix. Looking at the differences between the conditions regarding informative component, both the request (+9,7%, p<.001) and combination (+6,8%, p=.019) condition had a significant increase of informative elaborations from time 1 to time 3, only partly confirming H3c. In more detail, only the request condition (+6,5%, p=.022) increased significantly from time 1 to time 2. Looking at differences between the conditions, the combination condition had a significantly higher percentage at time 2 (30%) and at time 3 (35%), compared to the control (respectively 22%, p=.041 and 24%, p=.009) and content (respectively 22%, p=.038 and 25%, p=.034) condition.

These findings suggest that offering additional structure for the assessee and assessor in the PFB process has the potential to augment the percentage of informative comments. As such, assessors could be more inclined to provide informative elaborations, in which they motivate why certain criteria were met or not. This is an important finding, as previous research has stressed that offering justifications for a certain evaluation is essential in view of students' performance (Gielen et al., 2010; Walker, 2015), and that students should therefore be taught to offer justifications more frequently (Webb & Mastergeorge, 2003). Regarding suggestive elaborations, the control condition had the highest percentage at all times. At time 1, the control condition (35%) had a significantly higher percentage of suggestive elaborations, compared to the combination condition (25%, p=.004). At time 2, the control condition (34%) had a significantly higher percentage of suggestive comments compared to the request (24%, p=.003) and combination (25%, p=.005) condition, revealing a significant decrease from time 1 to time 2 for the request condition (-6,3%, p=.021). At time 3, no significant differences were found anymore between the conditions.

As it is important that assessors can indicate in their PFB messages how the assessee can improve their future work i.e. the feed forward component (Hattie & Timperley, 2007), results indicated that students provide slightly more suggesting than informative elaborations, independently from the structuring degree of the PFB template. These findings suggest that assessors, who only employ a structured PFB template, appear to be more inclined to offer suggestive comments in their PFB messages. Although research revealed that the presence of suggestions did not have a significant impact on performance improvement (Gielen et al., 2010), other research underlines that feedback should include improvement strategies (Duijnhouwer, Prins, & Stokking, 2012), in order to be beneficial for students' writing performance (eg. Stern & Solomon, 2006). When students were answering on a feedback request, they appeared to be more inclined to write justification to substantiate their previously made judgement.

In order to close the PFB loop, this study also examined whether students agreed or disagreed with the received PFB and secondly, if they actually implemented or simply ignored the received PFB. As illustrated in Figure 7, results reveal that all students mainly agreed on average with 60% of the received PFB messages, instead of disagreeing or partly agreeing,

confirming H4a. Even though students agreed at time 1 on average with 60% of the received PFB, results revealed that the control condition agreed significantly more (70%) at time 2, compared to the conditions who used the PFB request namely, the request (55%, p=.045) and combination (58% p=.107) condition, not confirming H4c. However, no significant differences were found anymore at time 3. Although results are rather fluctuating, this finding suggests that when only a PFB template is incorporated in the PFB process. From time 1 to time 3, it appeared that only the content condition (+15%, p=.011) had a significantly higher agreement percentage. This finding suggests that when PFB is underpinned with additional information obtained by the content checklist in order to justify particular comments, students will be more inclined to agree with the PFB over time.

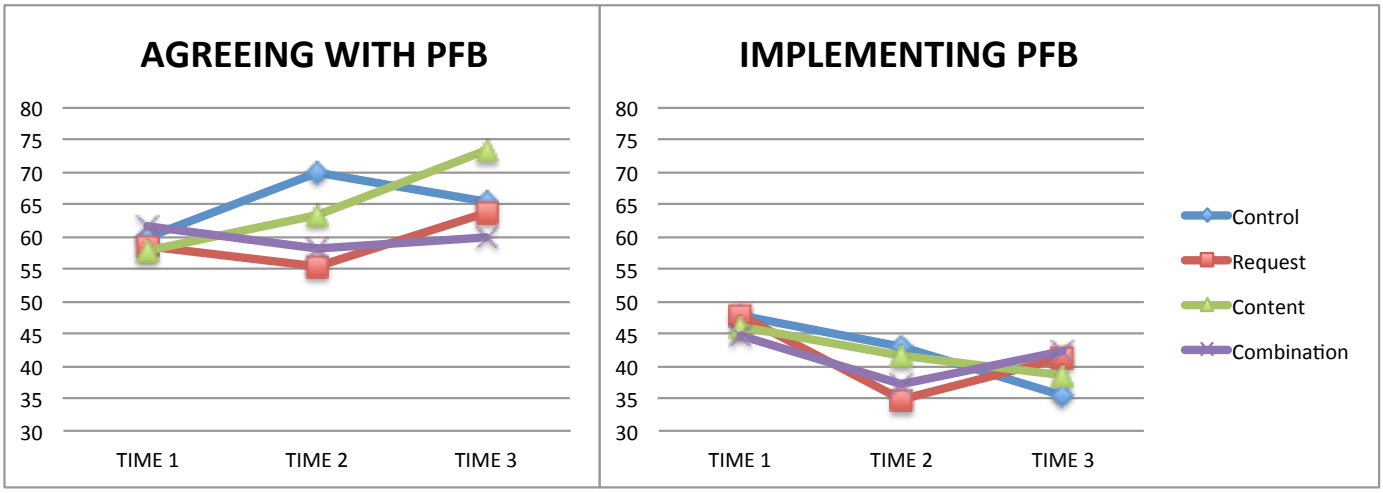


Figure 7. An overview of the percentage of agreement and implementation of the PFB

Inspired by research that underlines the need for examining short-term and long-term learning effects of PFB and the effects of an a posteriori reply form, this study aims to provide more insight on the actual implementation of the PFB, as feedback left unattended or not acted upon cannot be effective (Gielen, at al., 2010). The results indicated that students implemented an average of 46% of the received PFB, in order to adapt their draft version of the abstract at the start of the assignment, not confirming H4b. However, no significant differences were found between the conditions. In more detail, only the request condition (-13%, p=.015) implemented significantly fewer adaptations from time 1 to time 2, while all other conditions demonstrated a non-significant decline. Nevertheless, both conditions who employed a PFB request increased non-significantly from time 2 to time 3. Related to this, previous research claimed that students, who receive personalised PFB through 'feedback-on-demand', would be more inclined to apply the feedback (Gielen, et al., 2010). The fluctuating results suggest that additional structure for the assessor or assessee does not seem to have a real impact on the actual implementation, which not confirms H4d. Yet, results revealed a significant decline from time 1 to time 3 for the control condition (-12%, p=.021), which suggests that when the assessee and assessor receive no additional structure, besides the structured PFB template, these students would implement

fewer adaptations over time. As shown in the appendix, the subcategories 'disputed' and 'partly disputed' will not be reported, due to percentages less than 5%.

Conclusion

This study examined the impact of instructional interventions that further specify the role of the assessor and assessee on the specific PFB content, during multiple peer feedback cycles against the background of writing assignments in the first year higher education. We share the view that more elaborate and specific feedback, contributes positively to a peers' performance (eg. Hattie & Timperley, 2007). However, the findings of this study revealed that providing additional structure for the assessee and assessee did not have a significant impact on the actual PFB content. Taking into account the fluctuations over time, the percentages of the different categories did not reveal any significant differences between the conditions at time 1, 2, or 3. The highly structured stepwise PFB procedure, instigated by the instructor, could be a first reason to explain this outcome. Another explanation could be that the PFB template, which was used by all conditions at all times, provided already an appropriate level of structure, which caused assessors to formulate similar PFB content in their messages.

In general, results indicated that all students provided a balanced proportion of verifications and elaborations, with moderately more elaborations, suggesting that they had a general understanding of how to formulate effective feedback (Narciss, 2008). Following, all students habitually provided more positive verifications that are primarily focused on particular criteria, in which they confirm whether particular criteria were achieved or not. Also, results indicated a descending percentage of negative verifications over time for all conditions. Related to this, research stressed that providing justifications are more important than accurate critique in the form of negative comments (Gielen et al., 2010). Students require not only feedback that tells them if they dealt with particular criteria correctly or not, but also why and what they should do about it to improve (eg. Coll, Rochera, & De Gispert, 2014). In this respect, students provide habitually a rather balanced proportion of informative and suggestive elaborations that are predominantly focused on particular criteria. This is in line with a large body of research, which claims that feedback should not only inform the assessee why particular criteria were correct or not, but also include suggestions in their PFB messages on how future performance can be improved (e.g. Hattie & Gan, 2011). However, a request and/or content checklist appeared to be a useful instructional intervention to augment the informative component when students elaborate on prior judgements in their PFB messages. In this view, previous research claims that revisions in writing assignments are mostly triggered by justifications (eg. Walker, 2015).

In order to find out more on how students actually use feedback (Walker, 2015), this study gave the opportunity to the assessees to actually reflect on the received PFB, but also how they used it when revising their draft version (Boud, 2000). We found that firstly, students mostly agree with the feedback messages they receive and secondly, that students implement almost half of the feedback messages they receive. Remarkably over time, it appeared that students generally agree more with the feedback after more practice occasions, but they also implement

less of the received feedback. As the quality of writing performance increases for all students over time (Gielen & De Wever, 2015c), a reason for this could be that students receive more positive verifications and less suggestive comments after more practice occasions, as their work corresponds more with the expected performance. As a result, students could agree more with the comments they receive, but as well implement less adaptations over time.

One limitation of this study is that a random subsample of sixteen groups (four groups from each condition) out of 27 groups in total was selected for segmentation and coding. Due to work constraints, it was not feasible to include more groups for the data analysis. Therefore, findings of this study could be expanded and replicated with larger samples, more diverse student populations and a variety of courses. Another limitation is that this study did not examine the content of the PFB request or the content checklist into detail. Therefore, analyzing the specific content of the PFB request and the content checklist could be valuable in order to relate this content firstly to the assessor's given PFB and secondly to the assessees' corresponding PFB evaluation, could be an interesting direction for future research. In order to do so, this study proposes a content analysis scheme to analyse PFB messages and the evaluation of the feedback in different contexts, as a starting point for future experimental research. In function of feedback effectiveness (Narciss & Huth, 2004), we believe that instructional interventions such as a PFB request and a content checklist are useful approaches to increase the PFB content quality. Although no significant differences were found, we notice that especially after multiple practice occasions the informative elaboration component augments in students' PFB messages, which is essential in function of future performance (Walker, 2015). To conclude, the aim of this study was to find out how structuring the role of the assessee (asking and evaluating PFB) and the assessor (preparing and providing PFB) can have a beneficial impact on the peer feedback content quality, with the underlying purpose to increase the potential impact of PA and boost students' learning in higher education.

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Wiley, New York.

Baker, M. J., & Lund, K. (1996). Flexibly structuring the interaction in a CSCL environment. In P. Brna, A. Paiva & J. A. Self (Eds.), *European conference on artificial intelligence in education* (pp. 401–407). Lisbon, Portugal.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38*, 92-l13.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.

Boud, D., & Molloy, E. (2013). Rethinking Models of Feedback for Learning: The Challenge of Design. *Assessment & Evaluation in Higher Education, 38*, 698-712.

Butler, L. B., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*, 474-482.

Carless, D. (2015). *Excellence in University Assessment: Learning from Award-winning Practice*. London: Routledge.

Cheng, K. H., Liang, J. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *Internet and Higher Education, 25*, 78–84.

Cho, K. & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline. *Computers and Education, 48*, 409–426.

Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers and Education, 52*, 141–146.

Coll, C., Rochera, M.J., & De Gispert, I. (2014). Supporting online collaborative learning in small groups: teacher feedback on learning content, academic task and social participation. *Computers & Education, 75*, 53-64.

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction, 22*, 171–184.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66.

García, A. S., García-Álvarez, M. T., & Moreno, B. (2014). Analysis of assessment opportunities of learning spaces: On-line versus face-to-face methodologies. *Computers in Human Behavior, 30*, 372-377.

Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585-594.

Gielen, M., & De Wever, B. (2015a). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449.

Gielen, M., & De Wever, B. (2015b). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315-325.

Gielen, M., & De Wever, B. (2015c). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education, 88*, 370-386.

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.

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133-152.

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment and Evaluation in Higher Education, 34*, 105–114.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice, 41*, 33–39.

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

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* NE: Center for Instructional Innovation, University of Nebraska-Lincoln.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31,* 199–218.

O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In: A. M. O'Donnell, & A. King (eds.), *Cognitive perspectives on peer learning* (pp. 179-196). Mahwah, NJ: Erlbaum.

Prins, F. J., & Mainhard, M. T. (2009, August). *Fostering student's self-regulation during feedback dialogues in vocational education.* Paper presented at the 13th biennial conference of the European Association for Research on Learning and Instruction, Amsterdam, The Netherlands.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465-477.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12*, 8-22.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-44.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: the role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623-637.

Sambell, K., McDowell, L., & Montgomery C. (2013). *Assessment for Learning in Higher Education*. London: Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Sluijsmans, D., Brand-Gruwel, S., Van Merriënboer, J. J. G. & Bastiaens, T. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*, 23–42.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced*

multilevel modeling. London: Sage.

Stern, L. A., & Solomon, A. (2006). Effective faculty feedback: The road less travelled. *Assessing Writing, 11*, 22–41.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29-48.

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.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages.* Paper presented at the EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*, 20–27.

Tsai, Y. C., & Chuang, M. T. (2013). Fostering revision of argumentative writing through structured peer assessment. *Perceptual & Motor Skills, 116*, 210–221.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort but what is the effect? *Assessment and Evaluation in Higher Education, 35*, 277– 289.

Van-Dijk, D., Kluger A. N. (2004). Feedback sign effect on motivation: Is it moderated by regulatory focus? *Applied Psychology: An International Review, 53*, 113-135.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5–13.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*, 186–193.

Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education. 40*, 232-247.

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39*, 73–97.

Appendix

Table 1

Coding scheme for analysing peer feedback content quality (modification based on Strijbos, et al., 2012)

Category	Subcategory	Description	Examples
Peer feedback style	Verification	Is the feedback sentence an evaluative statement expressed as a positive or negative remark on past performance, based on an initial criteria or not?	*Your limitations are not included in the abstract.* *Well written!*
	Elaboration	Is the feedback sentence an informative statement that builds further on verification or remark expressed as e.g. a question, a confirmation, a suggestion or a justification?	*Your limitations are lacking, so please try to include them in your final version.* *I like it because you use your own words.*
	General	Is the feedback sentence a neutral statement, which doesn't have the characteristics of a verification or elaboration?	*This week, I'm providing feedback on your second abstract.*
Verification type	Positive	Is the feedback sentence a positive evaluative statement?	*The intention of the study is well formulated!*
	Negative	Is the feedback sentence a negative evaluative statement?	*I can't find your limitations in the draft!*
	Neutral	Is the feedback sentence a neutral evaluative statement?	*In your abstract, you refer to the methodology.*
Verification focus	Abstract general	Is the feedback sentence an evaluative statement that gives general details about the overall abstract, but without referring to particular criteria?	*All necessary components are included in your draft version.*
	Criteria general	Is the feedback sentence a general evaluative statement that provides minimal details about a particular criteria, or that merely expresses if a particular criteria is correct, present, or not?	*The problem statement and research purpose are present*
	Criteria specific	Is the feedback sentence an evaluative statement that provides profound specific details about the extent to which particular criteria were met in the past performance?	*The introduction summarises perfectly the intention of the research, by mentioning the research purpose before stating the actual context of the research.*
	Language	Is the feedback sentence an evaluative statement about language features such as verbs, translations, pronouns, spelling, grammar, sentence construction and layout?	*There are some little spelling mistakes in your conclusion.*
Elaboration type	Informative	Is the feedback sentence an informative statement, which gives more details about a previous evaluative statement without activating the student to adapt his work?	*Your intro is well formulated! (Pos. Verification) ... Particularly, I like how your abstract deals with the shift from the intention of the study towards the problem statement.*
	Suggestive	Is the feedback sentence a suggestive statement, which gives more details about a previous evaluative statement with the purpose to activating the student to adapt his work?	*In your final version, you should integrate the limitations, which you can find on page 9.*

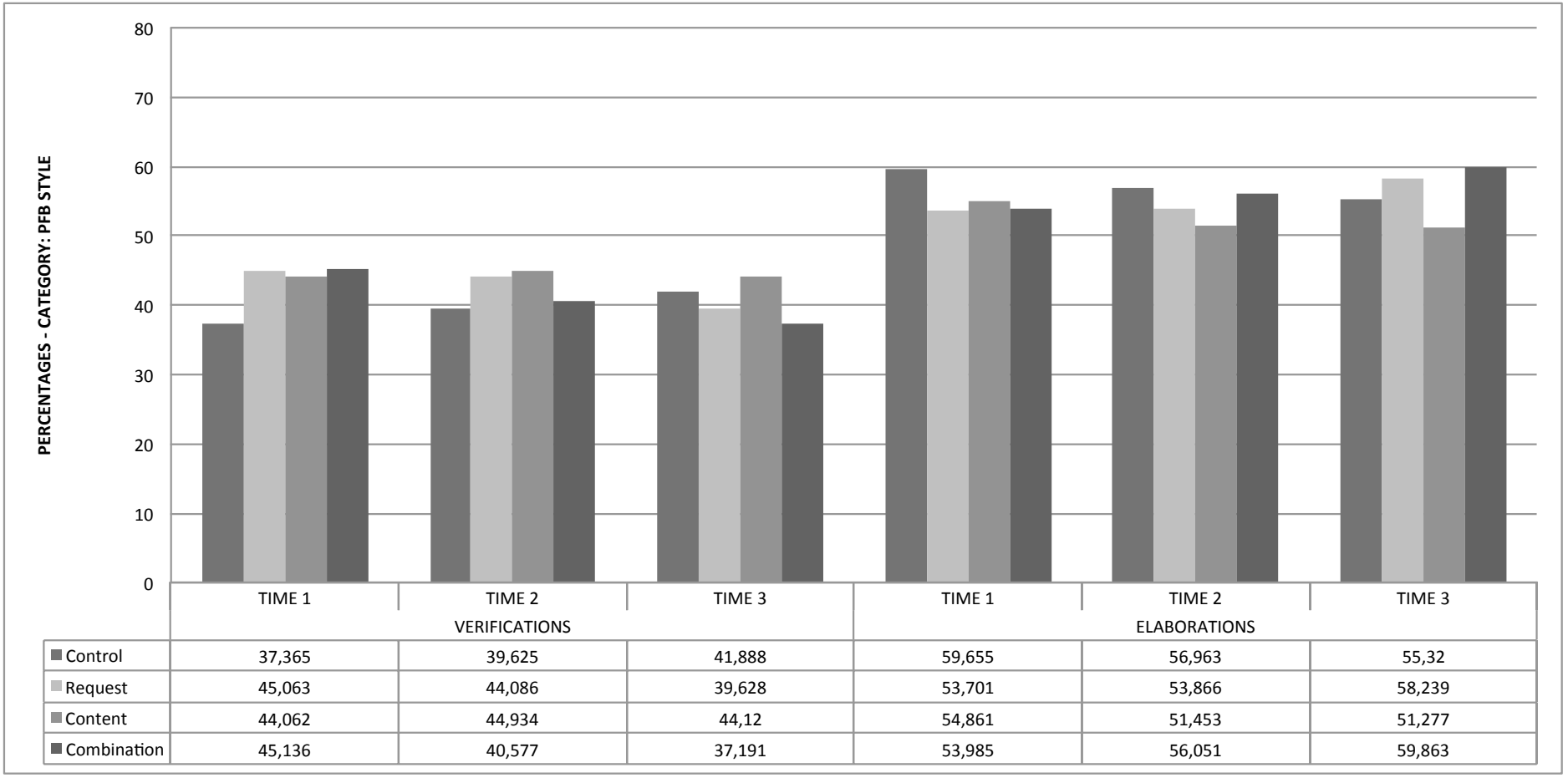
Elaboration focus	Abstract general	Is the feedback sentence an elaboration that gives general details about the overall abstract, but without referring to particular criteria?	*I believe you can still improve the quality of your abstract*
	Criteria general	Is the feedback sentence a general elaboration that provides minimal details about a particular criteria, or that merely expresses if a particular criteria is correct, present, or not?	*Maybe you should try to merge more the intention of the research and the problem statement*
	Criteria specific	Is the feedback sentence an elaborated that provides profound specific details about the extent to which particular criteria were met in the past performance?	*I would add the number of participants and more details about the context in the methodology section*
	Language	Is the feedback sentence an evaluative statement about language features such as verbs, translations, pronouns, spelling, grammar, sentence construction and layout?	*Once you finish, please check for spelling mistakes*
Evaluation agreement	Agree	Does the assessee states clearly that he agrees with the assessor when evaluating the received feedback?	*I believe your suggestion regarding lacking limitations is correct*
	Disagree	Does the assessee states clearly that he does not agree with the assessor when evaluating the received feedback?	*I believe that my original problem statement was already clear enough*
	Partly agree	Does the assessee state he partly agrees with the assessor when evaluating the received feedback?	*I followed your advice on the limitation, but I believe that the number of participants should be included as well*
Evaluation Implementation	Adapted	Does the assessee indicate he adapted his work based on the received feedback?	*Thanks a lot for your suggestion. It was very helpful to improve my final version*
	Ignored	Does the assessee indicate he ignored the received feedback?	*I did not*
	Disputed	Does the assessee dispute the received feedback?	*I do not agree with how you formulated the problem statement*
	Partly adapted	Does the assessee indicate he partly adapted his work based on the received feedback?	*I changed only a few of your language suggestion.*
	Partly disputed	Does the assessee partly dispute the received feedback?	*I did not implement all your suggestion regarding the methodology, but I like how you rephrased my initial idea*

PEER FEEDBACK REQUEST	
CRITERIA	**REMARKS / QUESTIONS / ISSUES?**
INTENTION	• Should I also include something about the specific focus of this study? • ... • ... • ...
PROBLEM STATEMENT	• ... • ... • ... • ...
METHODOLOGY	• I wonder if I explain this section profoundly enough. How should I make it more concise because I believe it is rather long at the moment • Is this part structured enough? • ... •
RESULTS	• Did I overlook any important findings? Extra things to include? • ...
CONCLUSION	• Does my conclusion cover the message of the whole research study? • ... • ...
LIMITATIONS	• I cannot really find any limitations • Do you have a suggestion? • ... • ...
GENERAL	• Did I explain the specific terms well enough according to you? • ... • ... • ...

Example of 'Peer Feedback Request'

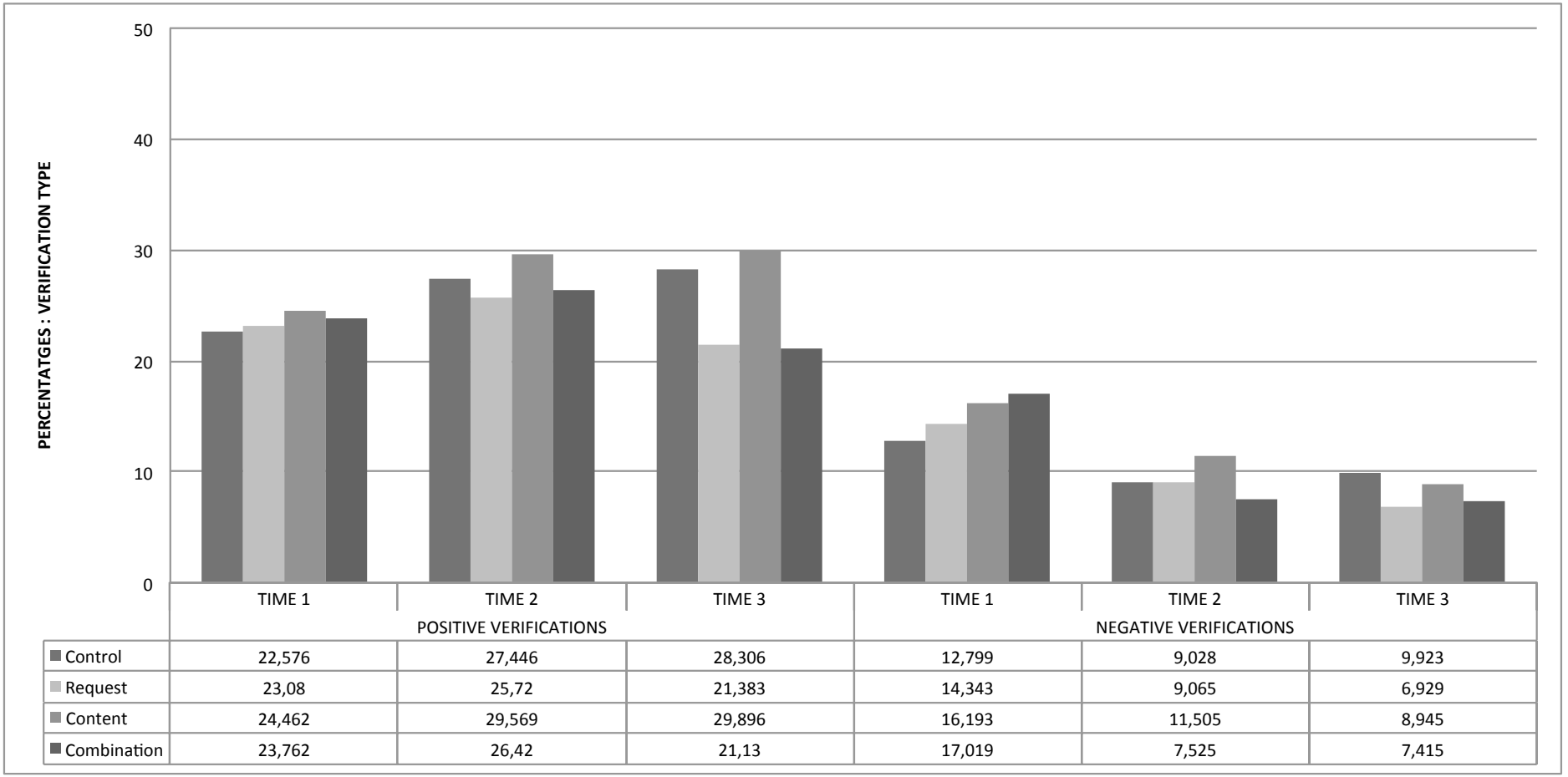
CONTENT CHECKLIST	
CRITERIA	**RELEVANT CONTENT FOR EACH CRITERIA**
INTENTION	• "...the goal of this paper is not so much to promote the use of commercial video games in education in se, as to understand, explain and predict changes in teachers' behavior in view of adopting these tools." • "The study contributes to an established body of research that has examined general reasons for playing video games ,the play behavior of teachers and teachers-in-training and teachers' acceptance of educational computer games." • "In this paper, a model-based approach to teachers' beliefs is presented and evaluated, based on the understanding that "teachers are faced with many variables that interact with each other to either facilitate or discourage the acceptance of technology"
PROBLEM STATEMENT	• "The present study focuses on the factors that influence the acceptance of commercial video games as learning tools in the classroom." • "When discussing teachers in relation to digital game-based learning, the focus is often on what they perceive as potential barriers to the implementation of games in their own practice." • "...to measure the concerns of the teachers regarding the difficulty of using games in their practice."
METHODOLOGY	• "...the focus is on teachers-in-practice." • "Secondary schools were contacted based on their denomination (i.e. community/subsidized public schools, and subsidized private schools), type of education (general, technical, and vocational) and geographical distribution." • "The teachers could fill in the questionnaires using the medium of their choice. This way, 505 teachers could be involved." • "The questionnaire consisted of three parts, examining demographic information, teacher related variables and the constructs of the research model." • "Demographic information included variables such as age and gender (0 female – 1 male). The teacher related variables included teaching experience (years), grade, and subject."
RESULTS	• "Most research variables are positively related. A notable exception is complexity, which relates negatively to personal innovativeness, behavioral intention, critical mass, experience and learning opportunities. While this negative relation was expected, considering the literature, the interrelations appear to be rather weak." • "High interrelations were found among behavioral

Example of 'Content Checklist'



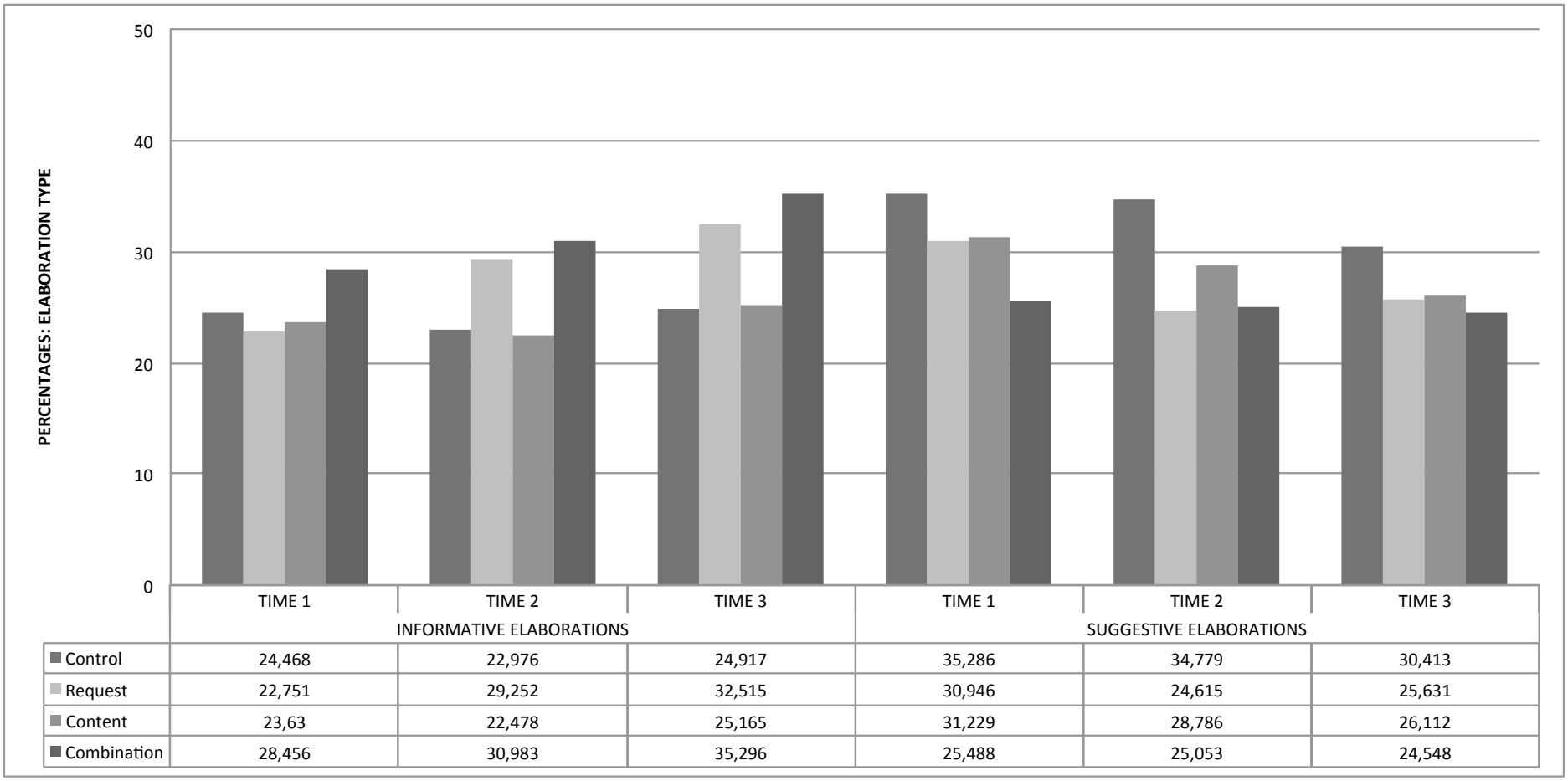
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
	VERIFICATIONS			ELABORATIONS		
Control	37,365	39,625	41,888	59,655	56,963	55,32
Request	45,063	44,086	39,628	53,701	53,866	58,239
Content	44,062	44,934	44,12	54,861	51,453	51,277
Combination	45,136	40,577	37,191	53,985	56,051	59,863

Percentages over time for the content category Peer Feedback Style



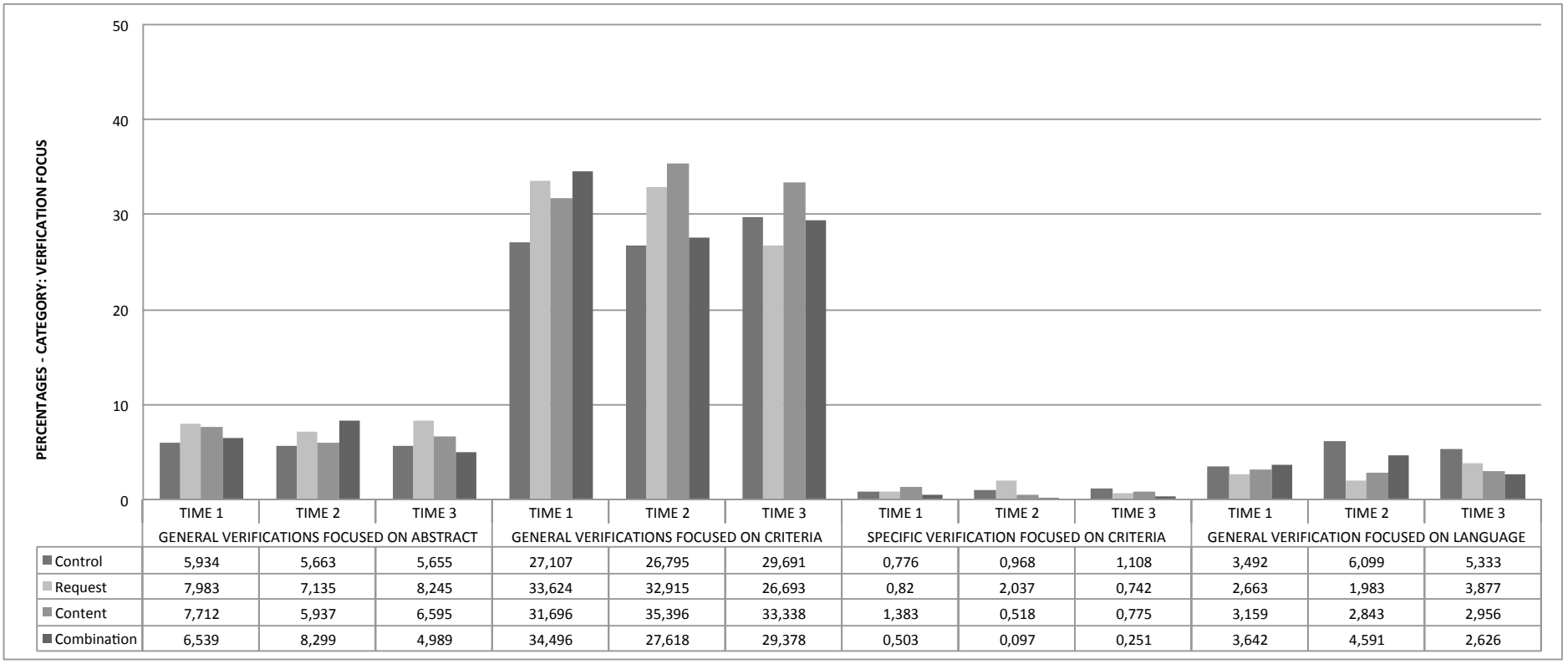
	POSITIVE VERIFICATIONS			NEGATIVE VERIFICATIONS		
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
Control	22,576	27,446	28,306	12,799	9,028	9,923
Request	23,08	25,72	21,383	14,343	9,065	6,929
Content	24,462	29,569	29,896	16,193	11,505	8,945
Combination	23,762	26,42	21,13	17,019	7,525	7,415

Percentages over time for the content category Verification Type



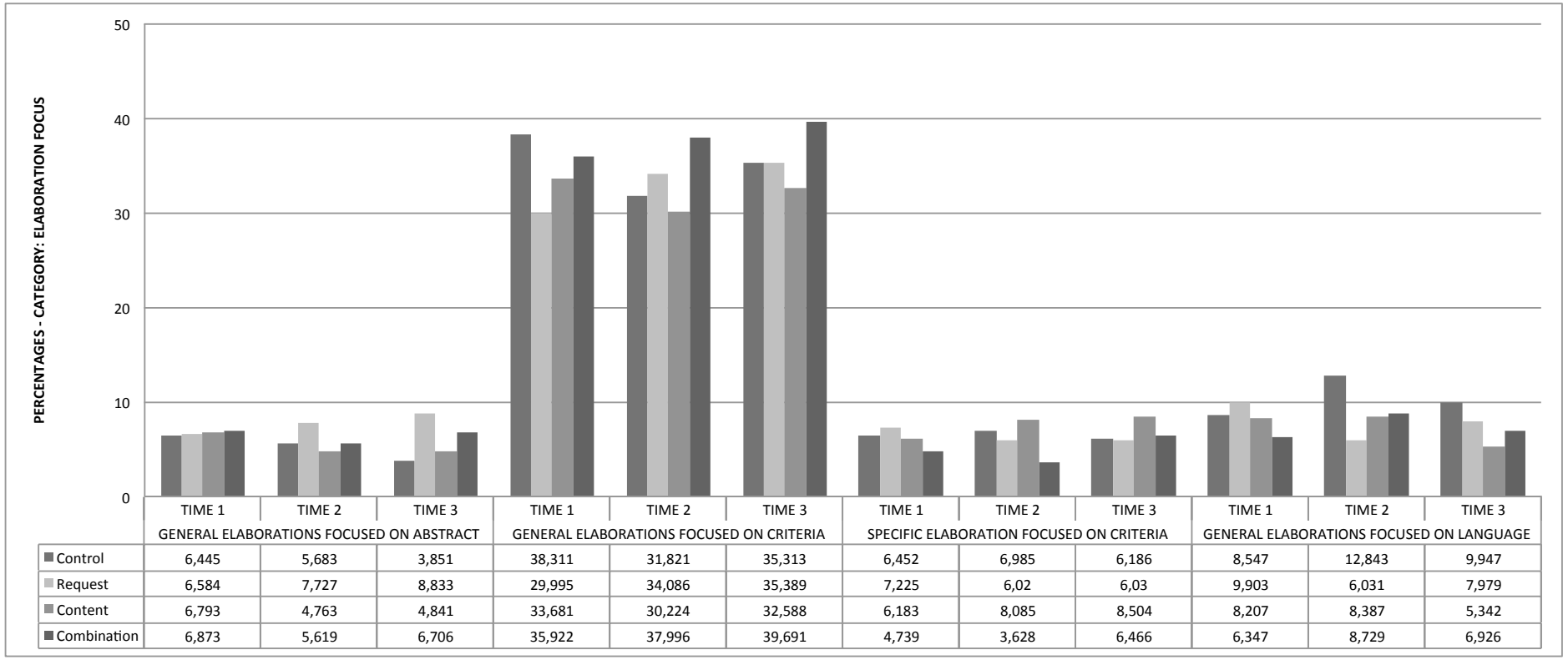
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
	INFORMATIVE ELABORATIONS			SUGGESTIVE ELABORATIONS		
Control	24,468	22,976	24,917	35,286	34,779	30,413
Request	22,751	29,252	32,515	30,946	24,615	25,631
Content	23,63	22,478	25,165	31,229	28,786	26,112
Combination	28,456	30,983	35,296	25,488	25,053	24,548

Percentages over time for the content category Elaboration Type



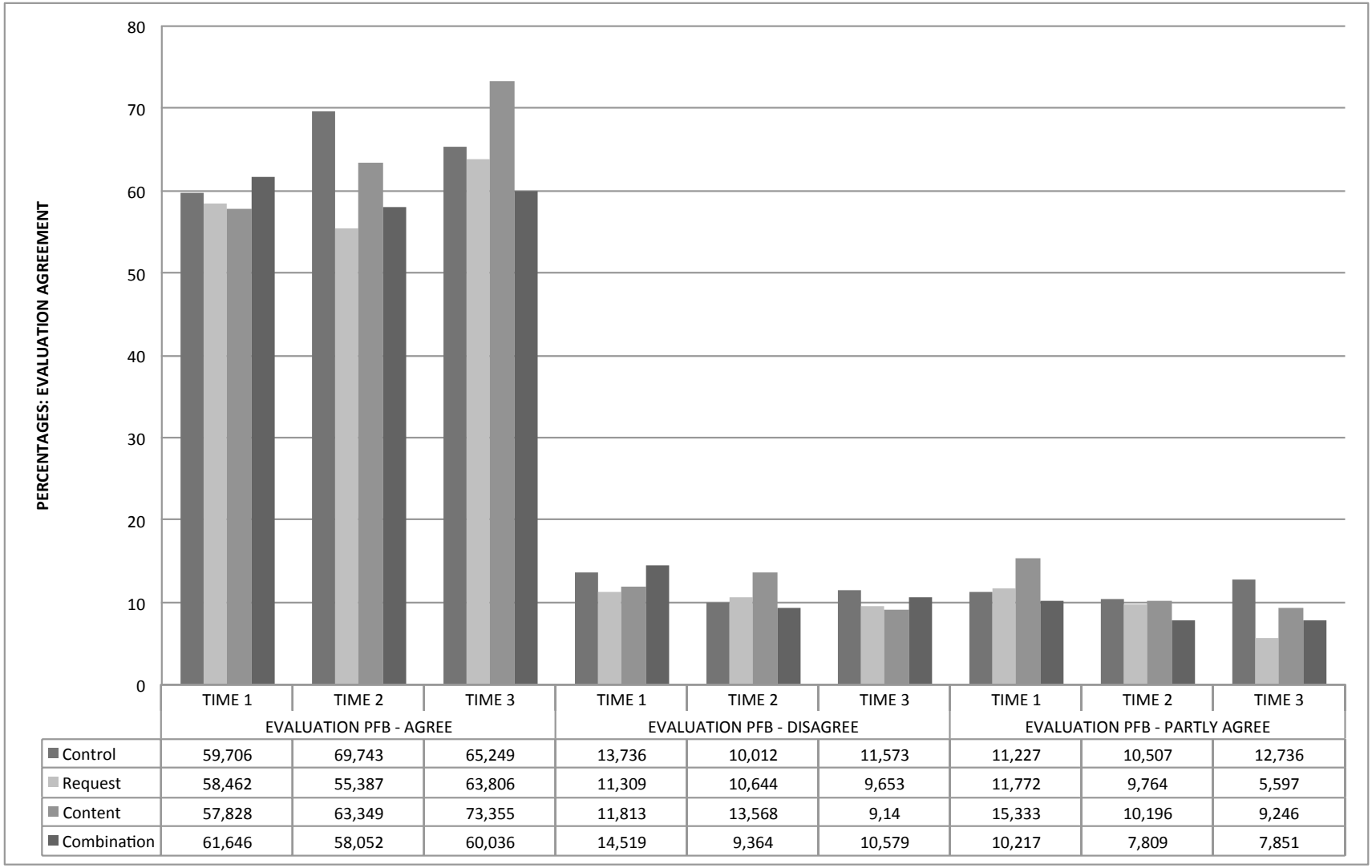
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
	GENERAL VERIFICATIONS FOCUSED ON ABSTRACT			GENERAL VERIFICATIONS FOCUSED ON CRITERIA			SPECIFIC VERIFICATION FOCUSED ON CRITERIA			GENERAL VERIFICATION FOCUSED ON LANGUAGE		
Control	5,934	5,663	5,655	27,107	26,795	29,691	0,776	0,968	1,108	3,492	6,099	5,333
Request	7,983	7,135	8,245	33,624	32,915	26,693	0,82	2,037	0,742	2,663	1,983	3,877
Content	7,712	5,937	6,595	31,696	35,396	33,338	1,383	0,518	0,775	3,159	2,843	2,956
Combination	6,539	8,299	4,989	34,496	27,618	29,378	0,503	0,097	0,251	3,642	4,591	2,626

Percentages over time for the content category Verification Focus



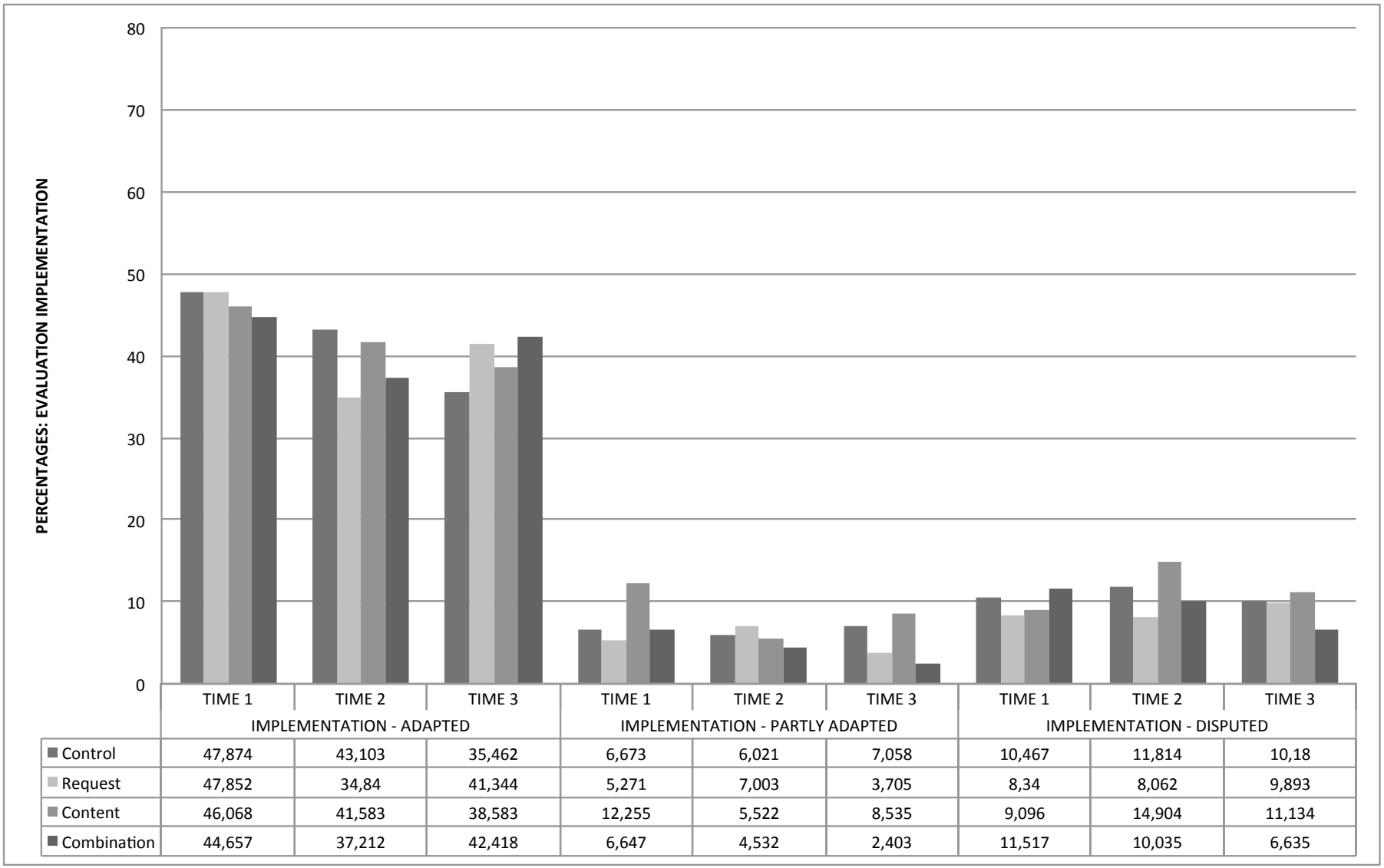
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
	GENERAL ELABORATIONS FOCUSED ON ABSTRACT			GENERAL ELABORATIONS FOCUSED ON CRITERIA			SPECIFIC ELABORATION FOCUSED ON CRITERIA			GENERAL ELABORATIONS FOCUSED ON LANGUAGE		
Control	6,445	5,683	3,851	38,311	31,821	35,313	6,452	6,985	6,186	8,547	12,843	9,947
Request	6,584	7,727	8,833	29,995	34,086	35,389	7,225	6,02	6,03	9,903	6,031	7,979
Content	6,793	4,763	4,841	33,681	30,224	32,588	6,183	8,085	8,504	8,207	8,387	5,342
Combination	6,873	5,619	6,706	35,922	37,996	39,691	4,739	3,628	6,466	6,347	8,729	6,926

Percentages over time for the content category Elaboration Focus



	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
	EVALUATION PFB - AGREE			EVALUATION PFB - DISAGREE			EVALUATION PFB - PARTLY AGREE		
Control	59,706	69,743	65,249	13,736	10,012	11,573	11,227	10,507	12,736
Request	58,462	55,387	63,806	11,309	10,644	9,653	11,772	9,764	5,597
Content	57,828	63,349	73,355	11,813	13,568	9,14	15,333	10,196	9,246
Combination	61,646	58,052	60,036	14,519	9,364	10,579	10,217	7,809	7,851

Percentages over time for the content category Evaluation Agreement



	IMPLEMENTATION - ADAPTED			IMPLEMENTATION - PARTLY ADAPTED			IMPLEMENTATION - DISPUTED		
	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3	TIME 1	TIME 2	TIME 3
Control	47,874	43,103	35,462	6,673	6,021	7,058	10,467	11,814	10,18
Request	47,852	34,84	41,344	5,271	7,003	3,705	8,34	8,062	9,893
Content	46,068	41,583	38,583	12,255	5,522	8,535	9,096	14,904	11,134
Combination	44,657	37,212	42,418	6,647	4,532	2,403	11,517	10,035	6,635

Percentages over time for the content category Evaluation Implementation

7 General conclusion and discussion

Chapter 7
General conclusion and discussion

Abstract

The research presented in this dissertation focuses on examining how peer assessment (PA) practices can be optimized in function of students' learning. More specifically, the aim of this dissertation is to investigate how structuring the role of the assessor and assessee in the PFB process can have an impact on the quality of students' writing performance and peer feedback (PFB). This final chapter provides a comprehensive discussion of the results obtained in the empirical studies, presented in chapter 2 to 6 and starts with an overview of the research objectives of this dissertation. Following, the main results of the studies are presented along with these research objectives. Based on the encountered limitations, future research aspirations are proposed. This dissertation concludes with implications for educational theory, practice and research.

Research objectives

The general aim of this dissertation was to gain more insight in ways to optimize PA practices in function of students' learning and more specifically in how structuring the role of the assessor and assessee in the PFB process can have an impact on the quality of students' performance and PFB. To deal with this main research aim and to tackle the different research challenges presented in chapter 1, three general research objectives are addressed:

> ***Research objective 1 (RO1):*** To explore the impact of different levels of structuring in the PFB process on students' performance
>
> ***Research objective 2 (RO2):*** To explore the impact of different levels of in the PFB process on the quality of students' PFB messages, measured by the Feedback Quality Index (Prins, Sluijsmans & Kirschner, 2006).
>
> ***Research objective 3 (RO3):*** To develop a content analysis scheme, which can be used to examine the specific content of students' PFB messages into more detail; and to implement this content analysis scheme to explore the impact of different levels of structuring in the PFB process on the content of students' PFB messages.

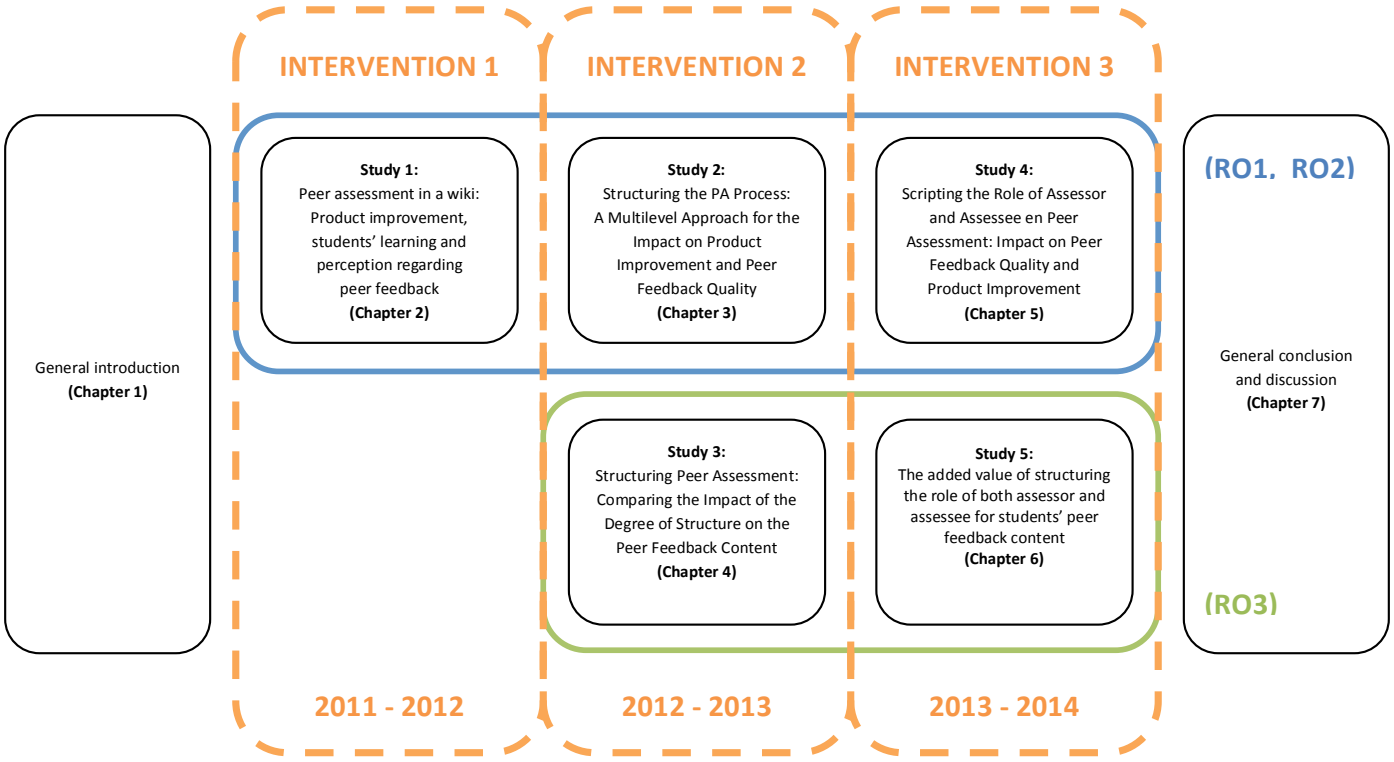


Figure 1. Overview of the 3 different interventions and 5 studies

Research objective 1 focuses on studying students' performance, while RO2 focuses on students' PFB quality as measured with the feedback quality index (FQI) of Prins, et al. (2006) on the other hand. Research objective 1 is dealt with in chapters 2, 3, and 5, respectively with respect to study 1, 2, and 3 as can be seen in Figure 1. Furthermore, research objective 2 is dealt with in chapter 3 and 5. Throughout the three studies, the way students are structured differed. In chapter two, the focus is on offering a basic structure in a PFB template versus no structure and differences in (1) students' learning, measured by multiple choice questions at the start and at the end of the assignment period, (2) the improvement of students' product from draft to final version and (3) students' perception towards the PA process in a CSCL environment in higher education, are discussed. In chapter 3, three different levels of structuring are implemented in the PFB template, namely: a no structure, a basic structure, and an elaborate structure condition. The differential effect of these different levels of structuring on (1) students' PFB quality and (2) product scores is discussed. Chapter 5 examines how an instructional intervention focusing on engaging both the assessor and assessee in the PFB process can be advantageous for the (1) quality of students' PFB and (2) written product. On one hand, the assessees are instructed to ask for feedback (PFB request) in the beginning and evaluate their received PFB at the end. On the other hand, the assessors are instructed to prepare the PFB content with the help of a content checklist, before formulating feedback on the provided PFB template. As chapter 5 builds further on the findings of chapter 3, all conditions incorporate the same optimized structured PFB template.

In function of RO3, chapters 4 and 6 deal with the development and implementation of a new content analysis scheme that is used to examine the specific content of students' PFB messages in a CSCL environment in higher education. In order to examine the specific content of the PFB messages that students provide to one another, chapter 4 reports on the proportion of the different PFB content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus, when a different low, medium or high level of structuring is implemented in the assessor's PFB template. Chapter 6 elaborates on chapter 4, as it examines whether the PFB content enhances when both the assessee (PFB request) and the assessor (content checklist) receive additional structure over time on top of providing and evaluating peer feedback with the help of a structured peer feedback template. Chapter 6 focuses on the same PFB content categories as chapter 4, but additionally it examines whether students actually agree with the received PFB and make adaptions accordingly. In order to do so, the categories (6) evaluation agreement and (7) evaluation implementation are added to the content analysis scheme.

In this concluding chapter, an overview of the main results of the different studies is presented along the three general research objectives. Table 1 gives an overview of how these three research objectives are tackled during the different chapters in the dissertation. Further, the main results that came to light in this dissertation are discussed. Additionally, limitations of this dissertation and directions for future research are presented. This chapter ends with suggestions for theory, research and educational practice.

	Chapter	Focus of the intervention studies	Research design and sample	Data-analysis techniques	RO
	CH1	General introduction (theoretical framework, research questions, research design and overview of the dissertation)			
INT.1	**CH2**	Study 1 investigates the added value of offering structure in the PA process. The aim of this chapter was to reveal differences as to (a) the learning effect, (b) the wiki product improvement and (c) students' perceptions of the PA process.	Quasi-experimental design *Sample:* Groups (n=38) Students (n=179)	Repeated measures ANOVA (SPSS) Repeated measures ANOVA (SPSS) Independent-samples t-tests (SPSS)	RO1
INTERVENTION2	**CH3**	Study 2 examines the impact of different levels of structuring in a PFB template for the assessor on (a) students' writing performance and (b) PFB quality.	Quasi-experimental design *Sample:* Groups (n=37) Students (n=168) Measurement occasions (n=3)	Multilevel analysis (MLwiN)	RO1 RO2
	CH4	Study 3 takes a closer look at the specific peer feedback content, when different levels of structuring are implemented in a PFB template in the PFB process of the assessor, regarding the occurrence of PFB content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus.	Quasi-experimental design *Sample:* Groups (n=9) Students (n=41) PFB forms (n=123) PFB segments (n=4717)	Content analysis AN(C)OVA (SPSS) MAN(C)OVA (SPSS)	RO3
INTERVENTION3	**CH5**	Study 4 investigates how an instructional intervention focused on engaging both the assessor (content checklist) and the assessee (PFB request) in the PFB process can have an impact on (a) students' writing performance (a) and (b) PFB quality.	Quasi-experimental design 2x2 factorial design *Sample:* Groups (n=27) Students (n=125) Measurement occasions (n=3)	Multilevel analysis (MLwiN)	RO1 RO2
	CH6	Study 5 takes a closer look at the specific peer feedback content, when both the assessor (content checklist) and the assessee (PFB request) in the PFB process receive additional structure over time, on top of providing and evaluating PFB with the help of a structured PFB template, regarding the occurrence of the following PFB content categories: (1) peer feedback style, (2) verification type, (3) verification focus, (4) elaboration type, (5) elaboration focus, (6) evaluation agreement, and finally (7) evaluation implementation	Quasi-experimental design 2x2 factorial design. *Sample:* Groups (n=16) Students (n=79) PFB forms (n=237) PFB segments (n=8440)	Content analysis Multilevel analysis (MLwiN)	RO3
	CH7	General conclusion and discussion (overview and discussion of the main results, limitations and suggestions for future research, and implications)			

Table 2. Overview of three different interventions throughout the dissertation

Main results

In function of the first research objective, the impact on students' performance is examined when different structuring degrees are implemented for the assessors and assessees in the PA process. First, study 1, outlined in chapter 2, is set up to examine the added value of offering structure in the PA process and this study involves two conditions: structured PFB (S-PFB) and non-structured (control). The aim of this chapter is to reveal differences in (1) students' learning, (2) wiki product improvement and (3) students' perception towards the PA process in a CSCL environment in higher education. The results presented in chapter 2 did not reveal significant differences in students' learning from pretest to posttest, nor between both conditions. As students were required to produce a draft and final version of their writing performance, the results show that the wiki product significantly improves for all students, when they have the opportunity to revise their initial draft before submitting the final result in the wiki-based CSCL-environment; this regardless of whether they receive additional structure in their PFB template or not. When students provide and receive peer feedback with the help of a structured PFB template, they have a stronger critical attitude towards the peer feedback, but they also perceive the received peer feedback as being more profound and detailed. However, the results demonstrate no significant additional impact on students' final product, when assessors are encouraged to employ a structured PFB template in the PA process.

Study 2, outlined in chapter 3, has been designed to examine the effect on students' writing performance, when a PFB template with different levels of structuring (no, basic or elaborate structure) is implemented. In both study 2 and study 3, students had to fulfill three consecutive assignments, in which they were required to participate in writing and assessment activities in order to compile abstracts based on scientific papers in the wiki environment. Multilevel analysis is performed to examine the effect of time (3 measurement occasions), student (N=168) and group (n=37) level variables on students' product scores. The results indicate an overall significant increase over time for all students. Related to this, the results demonstrate an overall significant increase of product scores of 9% from draft to final version. With respect to the provided level of structure in the PFB process, the only significant difference regarding product quality scores was found at time 3. After multiple practice occasions, it appeared that students of both the basic and elaborate structure conditions have significantly higher product scores, compared to students who did not receive additional structure in the PFB process. Otherwise, no other significant differences are found between the conditions. As such, study 2 stresses that offering additional structure in a PFB template, such as a few guiding questions or a bullet pointed list of criteria structured according to the three feedback principles feed up, feedback and feed forward (Hattie & Timperley, 2007), can be beneficial for the quality of students' writing performance.

Study 3, outlined in chapter 5, examines how an instructional intervention focusing on engaging both the assessor and assessee in the PFB process can be advantageous for the quality of students' written products. Every student (N=125) belongs to a group (n=27) of four to five

students. The main aim of chapter 5 is to examine the effect of structuring the role of the assessee and/or assessor by respectively providing them with a PFB request and/or content checklist, in combination with a structured PFB template. In order to do so, a 2x2 factorial design was set up, resulting in four conditions: (1) a control condition, (2) a feedback request condition, (3) a content checklist condition, and (4) a combination (feedback request + content checklist) condition. In function of RO1, these four conditions are being compared to reveal differences in product scores over time. The results of chapter 5 point out that from time 1 to 3, all students delivered writing performances of a significantly higher quality. From time 1 to time 2, both conditions that incorporated a PFB request (i.e. the one with and the one without the assessor completing the content checklist) demonstrates a significant incline in product scores. Since the product scores of the control condition also increased significantly from time 1 to time 2, this finding may question the actual necessity of providing additional structure or support in the PFB process in function of the product quality. We need to take into account that the PFB process has already structured to a certain extent since all conditions were using the same structured PFB template.

The second research objective of this dissertation is to identify differences in students' PFB quality, when assessors receive different levels of structuring in the PFB process. Described in chapter 3, study 2 examines the effect on students' PFB quality scores, measured by the Feedback Quality Index (Prins, et al., 2006). Multilevel analysis was performed to identify differences between the no structure, a basic structure, and an elaborate structure condition. As presented in chapter 3, the results reveal that the feedback scores increase significantly for all conditions from time 1 to time 3. With respect to the offered level of structure in the PA process, results of study 2 illustrate that students in the elaborate structure condition would have significantly higher feedback quality scores, if compared to students who merely received some guiding questions (basic structure condition) or who received no additional structure at all. However, the product scores increase significantly for all students over time.

Study 3, outlined in chapter 5, explores the added value to the quality of students' PFB when both the assessor and assessee are actively engaged in the PFB process. In this respect, four conditions are distinguished in a 2x2 factorial design. The first factor is whether the assessee is scripted or not scripted to ask for specific feedback on particular criteria and to evaluate the received PFB. The second factor is whether the assessor is scripted or not scripted to participate in a preparation task, before actually formulating feedback on a structured PFB template. When decomposing the increase over time, the results of Chapter 5 illustrate that all students formulate PFB messages of a higher quality over time. Although no dissimilarities had been expected between the conditions before the actual start of the intervention (as groups of students were randomly assigned to conditions), results do reveal significant differences in PFB quality scores between the four conditions at time 1. In view of RO2, we take a closer look at what happens with the quality of the PFB from time 2 onwards, when in the experimental conditions the assessee is required to ask for specific feedback and the assessor is required to complete a preparation task before the actual feedback provision. The results are rather inconclusive regarding the actual effectiveness of the content checklist for the PFB process.

Given that both conditions that implemented the content checklist had significant higher PFB quality scores at time 1, this could be a possible explanation for the non-significant increase over time. Completing a PFB request appeared to be a useful approach to increase the PFB quality, especially in the initial phase of performance, as both conditions had a significant increase in PFB quality scores from time 1 to time 2, with or without the assessor completing the content checklist. From time 2 to time 3, only the control condition showed a significant increase. As the significant differences in PFB quality scores disappear at time 2 and time 3, these findings suggest that students become more skilled assessors, when they have multiple occasions to practice.

Finally, the third research objective of this dissertation was the development and application of a content analysis scheme to examine the specific content of students' PFB messages during writing assignments in a CSCL environment in higher education. As an overarching ready-to-use content analysis scheme fitting these particular needs does not exist in the literature, this dissertation develops a new coding scheme for analyzing the content of PFB messages. This particular scheme is inspired by a recently developed coding scheme (Strijbos, Goozen & Prins, 2012), which is on its turn based on the generally accepted feedback framework (Narciss, 2008). In order to tackle RO3, we are particularly interested in finding out more about the specific feedback content students actually provide for each other. Study 2, outlined in Chapter 4, is set up to examine the impact on the PFB content quality, when assessors receive a different level of structuring in the PFB template. Through performing quantitative content analysis and analyses of (co)variance, Chapter 4 attempts to shed more light on possible differences between the no structure, basic structure, and elaborate structure conditions in particular, regarding the occurrence of PFB content categories: (1) PFB style, (2) verification type, (3) verification focus, (4) elaboration type, and (5) elaboration focus. Therefore, 9 of 37 groups (three groups of each condition) were randomly selected. All three feedback cycles are analyzed, which has resulted in 123 peer feedback forms of 41 students in total. After the segmentation and coding process, these forms result in a database of 4717 segments for content analysis. Regarding the PFB style, the results of Chapter 4 show that all conditions provide a reasonably balanced proportion of verifications and elaborations in their peer feedback messages. Adding a few guiding questions in the basic structure condition, appeared to significantly expand the proportion of elaborations in PFB messages, compared to students who receive no further structure.

Regarding verification type and focus, results show that all students offer mainly positive verifications and slightly more suggestive than informative elaborations, which are focused on the pre-defined criteria they need to for writing a good abstract. Interestingly, the results demonstrated that an elaborate structure in the PFB template has the potential to encourage students to formulate feedback on criteria level. When students receive no additional structure in the PFB template, they tend to formulate more general feedback on the overall assignment and on language aspects, in which they mainly confirm whether the performance corresponds with the expected outcome or not. This could be a possible explanation why students in the elaborate structure condition have significantly higher feedback quality scores, measured by the FQI, compared to students who merely received some guiding questions or who received no

additional structure at all. As such, it is recommendable for instructors to include a list of unambiguous criteria, whenever they offer a PFB template, in which student are instructed to apply this list of criteria according to the three feedback principles feed up, feedback and feed forward (Hattie & Timperley, 2007).

Lastly, study 3 examines the impact of instructional interventions that engage both the assessee (asking and evaluating feedback) and the assessor (preparing and providing feedback), on the specific PFB content during multiple peer feedback cycles against the background of writing assignments in the first year of higher education, which is outlined in chapter 6. For this reason, a 2x2 factorial design was set up to examine the effect of structuring the role of the assessee and/or assessor by respectively providing them with a PFB request and/or content checklist, together with a structured PFB template. Multilevel analysis was performed to indicate differences between the conditions regarding the percentages of the following PFB content categories: (1) PFB style, (2) verification type, (3) verification focus, (4) elaboration type, (5) elaboration focus, (6) evaluation agreement, and finally (7) evaluation implementation. The results of chapter 6 indicate that providing additional structure for the assessee and assessee do not profoundly affect the content of PFB messages. The only significant difference found between the conditions is that a request and/or content checklist appears to be a useful instructional intervention to augment the informative component, in which students elaborate on prior judgments in their PFB message, especially after more practice occasions. For the rest, no other significant differences in PFB content have been found. The results of study 3 reveal that all students provide a rather balanced proportion of verifications and elaborations, with moderately more elaborations. The verifications are mostly positive and focus on particular criteria. Besides, these students formulate a rather balanced proportion of informative and suggestive elaborations, predominantly focused on particular criteria. Remarkably, results indicate that all conditions offered less negative verifications in their PFB messages over time. A possible explanation can be that students simply require less feedback after more similar practice occasions, taken into account that the quality of the former writing product increased significantly for all students over time. In order to close the PFB loop (Boud, 2000), study 3 also examines whether students generally agree with their received feedback, and how they actually use the feedback when revising their draft version. In chapter 6, the results reveal that students mostly agree with the feedback and secondly, that students implement almost half of the feedback messages they receive.

General discussion

In this section, the most important results presented above are being discussed, with a focus on three general themes that occurred in this dissertation, i.e. '*Structuring versus practicing the PFB process*, '*Structuring the assessor's PFB template*', and '*Triggering peer feedback dialogue'.*

Structuring versus Practicing the PFB Process

In Chapter 1, we state that there is still no rigid approach on how to design PA practices in function of students' learning (Strijbos & Sluijsmans, 2010). In the literature, many questions remain unanswered on how PA should be implemented accurately into educational practice in order to boost students' learning in higher education (Sadler, 2010). As PA is an example of a complex task, we have to take into account that high-level cognitive processes that are required during PA practices, do not happen automatically. For this reason, previous research stresses the importance of further specifying the roles and activities for the learners involved during PA practices (Topping, 1998). *Why do we initially want to structure the PFB process?* Previous research has raised its concern regarding the low quality of PFB comments and refers to '*production deficiency*' and '*availability deficiency'* as two essential reasons why it is necessary for students to receive instructional support or practice in order to enrich the quality of their PFB messages (Prins, et al., 2006).

The first reason could be that students cannot deliver appropriate feedback because they fail to instinctively know which suitable strategy to follow (i.e., *production deficiency*). By prompting specific feedback activities, instructional support can ease production deficiency. In this manner, feedback instruments such as "performance scoring rubrics with criteria, or structured feedback forms that force feedback providers to ask reflective questions and give suggestions for improvement can be valuable instruments to increase the quality of the peer feedback" (Prins, et al., 2006, p. 300). With respect to the different levels of structuring in this PFB process, it becomes apparent in this dissertation that structuring the role assessor (with the help of a preparation task or a template to provide PFB template) and the role of the assessee (by letting them ask for and evaluate PFB) in the PFB process do not really have an important impact on the writing performance, the PFB quality (FQI), or the content of the PFB messages. Throughout the different studies, it appears that the quality of the writing performance and PFB messages augmented for all students, whether or not they had received additional structure in the PFB process.

However in study 1, we do notice that structure can have an impact on students' perception towards the PA process. The results indicate that adding a few guiding questions to the PFB template such as 'What do you like about the work of your peer?' and 'What would you change in the work of your peer?' have an influence on students' perception. More specifically, students appear to have a more critical attitude towards the received feedback. As assessees, students

need to "critically review the PFB they have received, decide which changes are necessary in order to improve their work and proceed by making those changes" (for detailed description see, Hovardas, Tsivitanidou, & Zacharia, 2014, p. 135). For this reason, study 3 includes a section 'Evaluation of the received PFB' in the structured PFB to find out more on how students actually use feedback (Walker, 2015). The results illustrate that students mostly agree with the received PFB and furthermore, they implement almost half of the received PFB comments. The results also point out that the additional structure in the PFB template made the students perceive the received PFB as being more profound and detailed. Li, Liu, & Steckelberg (2010) discovered that students acknowledge the value of PFB, but that they were not always satisfied about the quality of their received PFB. In that study, the lack of constructive and more detailed feedback was associated with poor quality feedback.

The second reason could be that students simply do not master the appropriate skill in order to produce high quality feedback (i.e., *availability deficiency*) (Flavell, 1976; Veenman, Kerseboom, & Imthorn, 2000). In order to ease availability deficiency, only instructional support will not be enough and hence, a methodological training will be more appropriate to acquire these particular feedback skills. Some research underlines that these complex skills require a considerable amount of practice (Sluijsmans, 2002), while other studies indicate that students do not really require training in assessment (e.g. Cho & MacArthur, 2010). In general, all studies illustrate that both the quality of the PFB messages and the actual writing product increase over time for all students. This is in line with research that claims that 'practice' leads to performance improvement (e.g. Kluger & Denisi, 1996) and that practice is essential to become a skilled peer assessor (e.g., Birenbaum, 1996; Van Steendam, Rijlaarsdam, Sercu, & Van den Berg, 2010).

More specifically, in study 1 and 2 it becomes apparent that the quality of the writing product increases significantly from draft to final version. In a culture of assessment for learning, research stresses that students should be given the opportunity to use the received feedback in order to improve their learning and achievement (Nicol & MacFarlane-Dick, 2004). Bearing in mind that the average effects of feedback are one of the highest in education, but also one of the most unpredictable in their influences (Hattie & Gan, 2011), we observed that quality of students' writing performance augment from draft to final after revising their initial performance. Furthermore, study 2 and 3 point out that practice in writing and assessment activities is essential for improvement. More specifically, the results point out that that student's product and PFB scores enhance significantly over time, when they have multiple occasions to practice. This is in line with research that underlines that students require practice in providing and receiving PFB in higher education (e.g. Nicol, 2010). For this reason, we share the view that it is important to take into account multiple practice occasions, when designing PA practices in function of students' learning.

Structuring the assessor's PFB template

In chapter 1, we stated that not all students would be able to act as skilled peer assessors from the start. Previous studies have already pointed out the need for structure and support to ensure effective feedback (e.g., Poverjuc, Brooks, & Wray, 2012). In order for the assessor's feedback to facilitate other peers' learning, the assessor not only needs to deeply process the assessees initial product, but also show planning and monitoring concerning how to formulate feedback in a way that a peer can benefit from it. Although instructional interventions that further specify the role of the assessor in the PFB process are more widely spread in the literature, a common approach on how to support the assessor during feedback provision is still lacking.

In this dissertation, we have developed a structured PFB template throughout the different studies in order to assist the assessor in formulating PFB messages in a structured way. In a first study, we added two guiding questions in the PFB template (What do you like? What would you change about a peer's work?), in which the first one focuses on providing feedback and the other one focuses on offering feed forward. Besides a list of predefined criteria, this also appears to have an impact on students' perception towards the received PFB. Although it appears that students are more critical, they do perceive the PFB messages as more profound and detailed. As such, these findings suggest that adding guiding questions can be beneficial for the PFB process. Since determining the accurate level of structuring appears to be the actual challenge (Dillenbourg, Järvelä, & Fischer, 2009), study 2 explores the effect when students receive different levels of structuring in the PFB template (i.e., no, basic and elaborate structure). As no previous studies investigated the impact of a similar peer feedback template with a different structuring level, this study attempts to answer what level of structuring is the most appropriate for the assessor during feedback provision (c.f. 'script granularity' concept of Kobbe, et al., 2007). In a final phase, we deliberately integrated the most valuable features from the previously tested PFB templates. Therefore, we combined these features into an optimized PFB template with a particular level of structuring, comprising four different sections presented in a table (See figure 2). The first section provides a list of criteria, to purposely stimulate students to consider each criterion separately. As feedback should be "on target, objective, focused, and clear" (Shute, 2008, p. 182), we support the point of view that feedback should be directed towards a list of (preferably) predefined criteria. In the second and third section, the assessor was structured to formulate feedback and feed forward: "How am I going?" (What progress is being made toward the goal?), and "Where to next?" (What activities need to be undertaken to make future improvement?) (Hattie & Timperley, 2007, p. 86). As such, the assessor will not only comment on how well the assessee is going, but also on the next steps to be taken (Topping, 2009). Based on the literature, the PFB template includes a fourth section, in which students were instructed to indicate, for each criterion and its corresponding feedback, whether the received PFB was valuable or rather irrelevant.

CRITERIA	FEEDBACK (PAST)	FEED FORWARD (FUTURE)	EVALUATION RECEIVED PFB
Criteria x			
Criteria x			
Criteria x			
Criteria x			
Criteria x			
Criteria x			
General remarks			

Figure 2. The developed structured peer feedback template

In general, the results in both studies reveal that the quality of both peer feedback and the final product increases for all conditions over time, but no real differences have been found between the conditions. A reason for this could be that all students, including the control condition, were using the same structured PFB template. Since the PFB process has already been structured to a certain extent, this finding implies that structuring the role of the assessor with the help of this optimized PFB template might be sufficient to enhance the quality of the assessor's PFB messages. Previous research underlines that especially the content of PFB messages is vital to the effectiveness of the feedback (e.g. Cho & MacArthur, 2010). In view of RO3, we have developed a content analysis scheme to analyse the quality of students' PFB messages in more detail in a CSCL environment in higher education. After examining the different content categories, it becomes apparent that the content of the PFB messages is rather similar between study 3 and 5, but also between the different conditions, no matter what different level of structuring they receive in the PFB template.

However, both studies point out that the PFB messages of all students habitually consist of a balanced proportion of verifications and elaborations. This is in line with research, which claims that successful feedback should include both verifications and elaborations (e.g., Bangert–Drowns et al., 1991; Mason & Bruning, 2001). In study 3, we notice that particularly students, who employed a PFB template comprising guiding questions or an elaborate structure, formulate more elaborations in their messages. Similarly in study 5, we find that students provide a rather balanced proportion of verifications and elaborations, but with slightly more

elaborations when all students receive the same level of structure in an optimized PFB template. A possible reason could be that the additional structure in the template encourages students to elaborate more on their PFB messages on the different aspects of a peer's performance when they are formulating feedback for their peers. Essentially, well-formulated feedback should provide an answer to three questions: 'Where am I going?' (feed up), 'How am I going?' (feed back), and finally 'Where to next?' (feed forward) (Hattie & Timperley, 2007). Traditionally, the feedback framework distinguishes between verifications and elaborations (Narciss, 2008). These results suggest that adding structure to a PFB template could increase the elaborative feedback component in students' PFB messages. Other research underlines that both verifications and elaborations define the quality of feedback messages (Kulhavy & Stock, 1981; Narciss, 2008). While verifications merely indicate how well particular criteria are achieved or not, elaborations provide more details on what was the reason why something is right or wrong and secondly, what students can do to enhance the quality of their performance (e.g., Hattie & Gan, 2011).

The results of both studies point out that students mainly formulate positive comments on particular aspects of the assignment, when students are engaged as peer assessors. This finding suggests that students are inclined to formulate more positive than negative comments. However, in study 3 we also notice that more structure in the PFB template results in more negative comments. When students use an optimized PFB template in study 5, our finding is confirmed that students formulated more negative comments in their PFB messages, in which they express that certain aspects of the assignment are not achieved. As such, the results of both studies suggest that adding structure to a PFB template can increase the negative feedback component in students' PFB messages. A possible reason for this is that a highly structured stepwise procedure encourages students to provide negative comments as well. In view of task revision, negative feedback may be necessary to expose shortcomings in a peer's performance, which on its turn can lead to increased effort (e.g. Bandura & Cervone, 1986). Without ignoring the importance of positive feedback, we support the view that students have to be confronted with the importance of formulating negative feedback as well. Previous research underlined that both positive and negative feedback could either increase or decrease performance (Kluger & Denisi, 1996). Despite this, the last study emphasizes that students still formulate less negative comments over time. A possible explanation for this could be that the product scores of all students improve after multiple practice occasions. As such, it might be the case that students' work simply requires less negative comments, when students have the opportunity to practice on similar writing activities.

Taking a closer look at the elaboration component, both studies point out that students' PFB messages consist out of a rather balanced proportion of informative and suggestive elaborations. In more detail, in study 3 we notice that students, no matter what level of structuring they receive in their PFB template, they provide slightly more suggestive comments. Furthermore, we also remark that when all students receive the same level of structure in an optimized PFB template in study 5, they all provide PFB messages with a more balanced proportion of informative and suggestive comments. These results suggest that adding structure to a PFB

template enables students to increase the percentage of informative elaborations in their PFB messages. A possible reason could be that the stepwise procedure of the PFB template encourages students to elaborate more in their PFB messages on the different aspects of a peer's performance by addressing the questions: 'How am I going?' (feed back), and finally 'Where to next?' (feed forward) (Hattie & Timperley, 2007). As such, a structured PFB template might trigger students to justify in their PFB messages why certain criteria were met or not, and to suggest how a peer's draft version can be improved, which in turn appears to be beneficial for its effectiveness (e.g. Butler, 1987). This is in line with research, which stresses that elaborations shall not only comprise what students can do to enhance the quality of their performance, but also indicate why something is right or wrong about particular aspects of the performance (e.g., Coll, Rochera, & De Gispert, 2014).

Regarding instructional interventions dealing with the role of assessor during feedback provision, the last study points out that an assessor's PFB template, which is enriched by the presence of a few guiding questions (basic structure) or a bullet-pointed list of criteria, structured according to the feedback principles (elaborate structure) of Hattie and Timperley (2007), can have an effect on the PFB quality. After implementing this developed PFB template in study 5, it becomes apparent that the content of the PFB comments generally remains rather equal at all practice occasions. A possible reason could be that the PFB template has already provided an appropriate level of structure, which causes assessors to formulate similar PFB messages. Nevertheless, we recommend to implement a structured PFB template in order to assist the assessor in the PFB process, as the results point out that students could be triggered to formulate PFB messages for all criteria separately, comprising more elaborations than verifications, more negative verifications and finally, more informative elaborations. For this reason, we believe that a structured PFB template can be beneficial for the quality of PFB messages in function of students' learning.

Triggering Peer Feedback Dialogue

In most studies on instructional interventions in PA practices, the instructional collaboration scenario is merely focused on the role of the assessor, while the role of the assessee is mostly forgotten (Gielen, et al., 2010). In an attempt to initiate feedback dialogue (Carless, 2015), in Chapter 1 we state that research stresses that we need to pay attention to the role of the assessee as well when designing PA practices (Fischer, Kollar, Stegmann, & Wecker, 2013). By engaging all actors intentionally in the PFB process, we strive to transform the feedback process into a more dialogic and elaborative process. In this dissertation, we have attempted to trigger a collaborative dialogue with two-way interaction between both the assessee and the assessor in the PFB process (Dippold, 2009), as further specifying the role of the assessor and assessee in function of a PFB dialogue could be advantageous for students' learning, as it enhances activity and interactivity (Topping, 1998). More specifically, the assessee participates in the PFB

dialogue by asking for specific PFB at the start, and eventually by evaluating the received PFB at the end of the PFB process. As such, PFB process can be perceived as "a communication process through which learners enter into dialogues related to performance and standards" (Carless, 2006, p. 280).

In general, the results of this dissertation indicate that after a few practice occasions all students manage to produce better writing products, but also formulate PFB comments of a higher quality, no matter what structure the assessor and assessee receive in the PFB process. Throughout the different studies, the results have suggested that different levels of structuring for the assessor and assessee do not have a major impact on students' writing performance, PFB quality or PFB content. A possible reason could be that the optimized structured PFB template, based on the findings of study 1 and 2, is sufficient for assessors to formulate their feedback in a highly structured stepwise PFB procedure and they do not require this feedback request to compose high-quality feedback. Also, providing unnecessary steps in order to prompt a PFB dialogue might unnecessarily complicate the PFB process and be another reason of obstructing students' learning. The former is in line with research that claims that excessively structured scripts can undermine students' learning (cf. 'over-scripting' - Dillenbourg, 2002). Moreover, it is possible that additional structure could be redundant when students have already gained experience in their role of assessor and assessee when multiple practice occasions are integrated in the instructional design. Even though the writing performances and provided PFB quality enhances for all students after multiple practice occasions, we will discuss why the role of the assessee deserves special attention in function of students' learning and why the assessee can be involved in the initial phase (by asking for feedback) and in the final phase (by evaluating the feedback) of the PFB process.

Although no real differences are found in the writing performance or PFB quality, results indicate that involving the assessees at the initial phase of the PFB process appears to be beneficial for the PFB content. More specifically, when assessees ask for specific feedback with the help of a PFB request, the results of study 3 suggest that assessors formulate significantly more informative elaborations in their PFB messages. This feedback component is particularly valuable for the assessee as it informs, evaluates, confirms or justifies previously made verifications on their performance. This is in line with previous research, which underlined that offering justifications for a certain evaluation is essential in view of students' performance (Gielen et al., 2010; Walker, 2015). As specific questions demand for more detailed and profound answers, a possible explanation is that the formulated questions guide the assessor's responses when formulating PFB comments. Similarly, it becomes clear in the first study that assessees will perceive the PFB comments of the assessor as more profound and detailed, when two guiding questions were added in the PFB template, referring to the principles of feedback and feed forward (Hattie & Timperley, 2007). Since the PFB comments are now in a direct response to the assessee's questions, a possible reason is that the assessor's comments are detailed with more informative elaborations, which address particular aspects of the assignment. This is in line with research that claims that the implementation of 'feedback on demand' in the PFB process (Gielen, et al., 2010) could motivate and direct the assessor to provide more 'responsive'

feedback (Webb, 1991), and thus provide PFB of a higher quality. Related to this, previous research showed that response-specific feedback appeared to augment learning efficiency (Shute, Hansen, & Almond, 2007). As more specific and elaborated feedback stimulates better performance and outcomes (Strijbos et al., 2010), we believe that letting assessees ask for feedback at the start of the PFB process, as part of the PFB dialogue, can be a valuable instructional intervention to increase the quality of students' PFB messages in function of students' learning.

In order to find out more on how students actually use feedback (Walker, 2015), the final study gives the assessee the opportunity to reflect on the received PFB. When closing the PFB loop (Boud, 2000), the results indicate that students predominantly agree with the feedback messages they receive and secondly, that students implement almost half of the feedback messages they receive. The results indicate that all students receive less negative verifications after more practice occasions. As the quality of writing performance generally increases over time and consequently corresponds more with the expected performance, this could be an explanation why students generally agree more with the feedback after more practice occasions, but they also implement less of the received feedback. Previous research stresses that PFB messages can only facilitate learning when the assessee actually takes up the feedback (Van der Pol, et al., 2008). However, feedback reception and revision do not automatically mean that learning will take place (Kollar & Fischer, 2010). However, other research highlighted problems concerning the depth, accuracy and credibility of the PFB messages, as the feedback comments, which they received from same-level peers instead of an instructor, might not be entirely relevant and accurate (Gielen, et al., 2010). In this respect, the results of study 1 show that assessees will be more critical towards the received PFB, when the assessor responds to two guiding questions that are offered in the PFB template. In view of triggering PFB dialogue, structuring the role of the assessee to assess the received PFB in the final phase of the PFB process could be beneficial for (1) the assessor, (2) the assessee, and (3) the instructor. More specifically, the assessee will be able to express if she or he agrees with the received PFB, finds the comments relevant and accurate, and indicate if he actually uses the feedback during task revision. On the other hand, it is rewarding for the assessor as this approach gives an indication whether the given feedback is appreciated and actually implemented by the assessee. For the instructor, the evaluation of the feedback can be consulted in any case of dispute during or after the assignment period.

Until now, it has become clear that in most of the intervention studies, the role of the assessee is limited to merely employing received PFB comments in a one-way direction for task revision (Kollar & Fischer, 2010). For this reason, we share the view that the PFB process shall be conceptualized as a dialogical and contingent two-way process, in which the active engagement of assessors and assessees is compulsory for a coordinated interaction between peers (e.g. Kollar & Fischer, 2010; Carless, 2015). In order to encourage a so-called feedback dialogue, we believe in the importance of structuring the role of the assessee firstly. This is accomplished in the initial phase by '*opening'* the feedback loop, in which the assessee is structured to ask for specific feedback and secondly, in the final phase by '*closing'* the feedback loop, in which

the assessee is structured to evaluate the received PFB comments. These are regarded upon as valuable instructional interventions in the PFB process in function of students' learning.

Limitations and directions for future research

Some limitations are inherently related to the studies presented in the different dissertation chapters. In each of the studies presented in chapter 2 to 6, specific study limitations are already addressed. As each chapter includes the limitations of the study in question, the limitations presented in this section deal with the overall picture of this research project. Based on the limitations and findings of the different studies in this dissertation, we also present some directions for future research.

Scope of this research

The main aim of this dissertation is to investigate how peer assessment (PA) practices can be optimized in function of students' learning. More specifically, the aim of this dissertation is to examine how structuring the role of the assessor and assessee in the PFB process can have an impact on the quality of students' peer feedback (PFB) and writing performance. Although the different studies are shedding light on both aspects, the relationship between the quality of the received feedback and the progress of task revision is underexposed. The studies in this dissertation do not investigate whether PFB has an impact on performance compared to a control condition without PFB (e.g., Gielen et al., 2010), or whether the quality of the PFB affects the performance improvement in order to examine its effect on learning. Related to the PFB quality, future studies can examine whether feedback of a higher quality, e.g. as measured by the FQI, results in a better task revision. Related to the PFB content, future studies can also investigate the relation between the specific content of PFB messages and the progress of task revision, with the help of the developed content analysis scheme in this dissertation. Eventually, future studies could discover which aspects of PFB content are particularly valuable for task revision. As such, studies could examine for example whether feedback messages, containing more informative elaborations, are more beneficial to task revision. Another limitation is that within study 3, we do not examine the content of the PFB request or the content checklist into detail. Analysing the specific content of the PFB request and the content checklist could be valuable in order to relate this content firstly to the assessor's given PFB and secondly to the assessee's corresponding PFB evaluation. In order to do so, the content analysis scheme presented in this dissertation can be adapted to analyse the content of a PFB request and content checklist in different contexts.

Population and sample

All studies included in the dissertation relate to first-year higher education students of an Educational Sciences program at university, predominantly females enrolled in the course 'Instructional Sciences'. Therefore, it is recommended to substantiate the ecological validity of the obtained results by replicating these studies with larger samples and increasing the number of intervention studies dealing with secondary and master students. And with groups of people who are younger, older, have larger gender diversity and other more diverse student populations from different courses in various programs. A second concern regarding the generalizability of the results relates to the small sample sizes, included in the studies reported in chapter 2, 3, and 5. Participants were randomly assigned to groups of about 4 to 5 students to work on a wiki task in a CSCL environment. A smaller sample size can lead to confounding of levels and increases the risk of less accurate estimates and standard errors (Hox, 2002). Statistical tests normally require a larger sample size to ensure a representative distribution of the population and to be considered representative of groups of people to whom results will be generalized or transferred. Future research could use larger samples to elaborate on our results and investigate whether similar estimates can be found.

Similarly, some questions might arise regarding the generalizability of the results obtained by content analysis of students PFB messages and reported in chapter 4 and 6. Only a few randomly selected groups of each condition are used for content analysis, in which multilevel analysis attempts to shed more light on the composition of the actual PFB content. As such, the data of study 2 consist out of 9 randomly selected groups out of 38 in total, while in study 3 only 16 out of 27 in total were selected for content analysis. This random selection of groups is necessary to be able to manually code all the PFB messages. While the amount of coded messages is definitely adequate to ensure reliable results, a selection of groups leads to some reduction of the full amount of available data. After segmentation and coding, the data for both studies result respectively in two substantial databases of 4717 segments for study 2 and 8440 segments for study 3, which are used to examine the specific content of the PFB messages. Due to time and financial constraints, it is not feasible to include more groups for data analysis. Yet, we believe that the findings of these studies can be expanded and replicated with larger samples. In addition, future studies can focus on studying the feasibility of some automatic coding. In this respect, Natural Language Processing (NLP) technologies appear to be very promising for text analysis.

Data-analysis

In general, the collected data in the present dissertation is predominantly quantified and then analyzed statistically in an objective manner (Creswell & Garrett, 2008). However, as the use of qualitative research is more encouraged in educational research, future research shall strive to collect, analyse, and interpret both qualitative and quantitative data (e.g., by means of semi-

structured interviews, focus groups, etc.) in a single study or a series of studies (Leech & Onwuegbuzie, 2009) and validate the obtained dissertation results. With the help of the developed content analysis scheme, we have a great opportunity to apply this scheme in the study 3 and 5. As such, we could statistically examine the PFB messages into more detail. However, we do need to consider that applying this content analysis scheme is a reduction of reality. Therefore, we believe that conducting a profound and detailed qualitative analysis of the PFB content can be valuable to gain more insight on various feedback quality features, such as the correctness of the feedback, which has currently not yet been explored.

Results

The results of this dissertation are conditional upon certain choices in the research design and procedure. Performance measures related to product and PFB quality are based on scoring rubrics for writing and assessing academic abstracts in a wiki-based CSCL environment. During all studies in this dissertation, students are not only engaged in the PFB process as peer assessors who are required to provide PFB, but also as peer assessees who receive PFB messages. Like this, we do not treat providing and receiving PFB as two isolated activities in this dissertation. In this respect, we cannot neglect that a significant increase of performance might not necessarily be the result of the received peer feedback. As PFB provision can be considered as a learning experience on its own, this could lead to writing improvement as well. Subsequently, it is difficult to predict whether the increase of the scores, measured by the scoring rubrics, is dependent on their involvement as assessor or as assessee. In order to shed more light on this, future research could examine the impact on writing performance, when students only participate as an assessee or assessee in the PFB process.

Implications

Implications for theory and empirical research

All studies throughout the dissertation are inspired by theoretical insights from different, but related research areas in the field such as assessment in higher education (e.g. Evans, 2013), peer assessment (e.g. Topping, 2009), peer feedback (e.g. Strijbos, et al., 2012), and collaboration scripting (e.g. Fischer, et al., 2013). In turn, this dissertation also contributes to these theories and their related empirical base in some important ways. Researchers from different theoretical orientations or who are rooted in different educational research fields (e.g. blended learning, instructional design, etc.) can therefore take advantage of the proposed studies and results. As such, we aspire that researchers are incited to tailor on the PFB process in function of students' learning.

In response to the lacking approach on how to effectively implement PA practices in function of students' learning, this dissertation examines the impact on students' PFB quality and product, when in the PFB process the role of the assessor and assessee is structured during writing and assessment activities in order to engage both actors more actively to trigger PFB dialogue. In this dissertation, we focus on one hand on instructional interventions, which engage the assessor such as instructing them to prepare the feedback through a content checklist and composing feedback with the help of a structured PFB template on the one hand. On the other hand, we focus on instructional interventions, which deliberately engage the assessee more actively in the PFB process by asking for feedback with the help of a PFB request, and evaluating the received feedback in the last section of the assessor's PFB template. Like this, we are convinced that these instructional interventions, which attempt to tailor the PFB process for both the assessor and assessee, extend the existing research body, as their empirical investigation is under addressed in assessment and scripting literature.

In this respect, the obtained results in this dissertation corroborate and extent previous research by showing that instructional interventions, in which the role of the assessor and assessee is structured to a certain extent in the PFB process, do not have a major impact on the quality of students' peer feedback (PFB) and writing performance. Nevertheless, it appears that these instructional interventions could be an effective way to prompt PFB dialogue (Carless, 2015). Involving the assessee actively at the start and at the end of the PFB process is preferable because it could trigger two-way communication. In this respect, we support the view that effective dialogue "should be *adaptive*, that is, contingent on students' needs; it should be *discursive*, rich in two-way communicative exchanges; it should be *interactive*, linked to actions related to a task goal; and last but not least *reflective* as it should encourage students and teachers to reflect on the 'goal–action– feedback cycle' (Laurillard, 2002, p. 503). Furthermore and very importantly, it becomes apparent that incorporating multiple practice occasions in the collaboration script is beneficial to students' PFB quality and product scores. This illustrates that

a well-considered instructional design for PA practices can help students to progress in their evolution towards more competent peer assessors. In this way, this dissertation can lead to a continual renewed interest in interventions in PA practices as a strategy to induce a larger repertoire of instructional interventions that optimize the PFB process in function of students' learning.

In particular, further investigation is required to unravel the relationship between the assessor's provided PFB and the assessee's performance improvement. It is therefore hoped that researchers are inspired by some crucial aspects that are taken into account in the intervention studies developed in this dissertation (chapter 2 to 6), and secondly that they are encouraged to engage in similar intervention studies. In this respect, researchers are incited to invest in similar intervention research and to go beyond by undertaking interdisciplinary mixed-method research, as described in the limitation section of this dissertation. In order to examine the PFB content quality in more detail, this dissertation (chapter 4 and 6) proposes a recently developed coding scheme (Gielen & De Wever, 2015) that is inspired by a coding scheme to analyse PFB messages (Strijbos, et al., 2012), which in turn built on the feedback framework of Narciss (2008). This coding scheme takes a closer look at the quality of peer feedback content, by concentrating on the peer feedback style, type, and focus in study 2. In addition, study 3 also takes into account students' evaluation of the received PFB, by looking at their agreement with and implementation of the feedback. In study 2 and 3, the results demonstrate that structuring the role of the assessee (e.g. asking and evaluating PFB) and the assessor (e.g. preparing and providing PFB) can have a beneficial impact on the peer feedback content quality, with the underlying purpose to increase the potential impact of PA and to boost students' learning in higher education. Therefore, we aspire that researchers will consider the developed content analysis scheme as a stepping-stone by which they are encouraged to engage in similar instructional interventions in the PFB process.

Implications for practice and policy

As to the implications for practice, the different studies in this dissertation, demonstrate that all students are able to provide PFB messages and compile writing performances of a significant higher quality after multiple practice occasions, no matter what level of structuring they receive in the PFB process. In this matter, study 2 till 5 provide means for instructors in practice to include several practice occasions into the collaboration script to augment the quality of students' feedback and performance. Therefore, this dissertation recommends instructional designers to integrate multiple practice occasions or feedback cycles, when designing collaborative learning and assessment activities.

Furthermore, the results illustrate that the developed and structured PFB template includes four different sections that provide a mean for instructors to implement this easy-to-use and

straightforward tool in PA practices. In the different studies, we have found that the first section, which comprised of a list of predetermined criteria, triggers students to provide PFB comments that are focused on particular criteria, which turns out to be a requirement to provide effective feedback (Shute, 2008). In the second and third section, students are required to provide feedback "How am I going?" and feed forward "Where to next?" in two different columns, based on Hattie and Timperley (2007). The obtained results illustrate that these two sections in the PFB template generated PFB messages consisting of a balanced proportion of verifications and verifications, which is a feature of successful feedback (e.g., Bangert–Drowns et al., 1991; Mason & Bruning, 2001). As such, instructors are able structure the PFB process in PA practices in a way that students not only verify whether another peer's past performance complies with particular criteria, but also that they are structured to offer valuable feed forward in function of future improvement, which is a key aspect of formative assessment (Sadler, 1989). In the last section, students have been required to evaluate the PFB comments for each criterion separately. The results indicate that students predominantly agree with the feedback messages they receive and secondly, that students implement almost half of the feedback messages they receive. As it is important to close the feedback loop in the PFB process (Boud, 2000), instructors could structure the assessee's role to evaluate their received PFB comments during PA practices.

In the assessment and scripting literature, it becomes apparent that a greater part of the studies deal with instructional interventions, which are merely focused on the role of the assessor, while the role of the assessee has often been forgotten in the collaboration script. In order to solve one-way feedback interaction (Nicol, 2010), this dissertation recommends instructors and instructional designers to optimize the PFB process during PA practices, and deliberately involve both actors to trigger a collaborative dialogue with two-way interaction (Dippold, 2009).

As to the implications for policy, we aspire that practitioners and policy-makers in higher education institutions in different countries, as well as for educational development and institutional research practitioners, are willing to confront the challenge of assessment for learning and think differently about it (Boud & Falchikov, 2007). In view of rethinking assessment in higher education, we share the view that PA is an important part of a shift towards more participatory forms of learning in our schools and universities (Kollar & Fischer, 2010). New, easily implementable instructional interventions that structure the role of the assessor and assessee in order to optimize the PFB process seem highly promising to facilitate PA practices in function of students' learning.

Final conclusion

In the literature, it became apparent that more research is necessary on optimizing the PFB process during PA practices in function of students' learning. As PA is an example of a complex skill, we share the view that collaborative learning and assessment activities require well-considered collaboration scripts during PA practices in order to contribute to students' learning and performance. Although the research presented in this dissertation covers only a tip of the iceberg concerning instructional interventions to optimize the PFB process, it provides a constructive starting point for future research.

In this dissertation, the results illustrate that the quality of both the writing performance and PFB messages increased significantly over time for all students, as long as they have the opportunity to practice on similar writing and assessment activities. These results corroborate and extent previous research by showing that instructional interventions, in which the role of the assessor and assessee receive a different level of structuring extent in the PFB process, do not have a major impact on the quality of students' peer feedback and writing performance. By developing and implementing a content analysis scheme, we have gained more insight in the actual composition of the PFB messages. While a structured PFB template guides the assessor through the PFB process in a stepwise manner, when preparing and formulating feedback and feed forward focused on criteriumlevel, this dissertation focuses as well on the role of the assessee, which is mostly forgotten in the literature. Therefore, in the last intervention the assessee is engaged from the start by 'opening' the feedback loop, in which the assessee is structured to ask for specific feedback and secondly, at the end by 'closing' the feedback loop, in which the assessee is structured to evaluate the received PFB comments. In view of triggering a so-called PFB dialogue (Carless, 2015), we support the view that new and easily implementable instructional interventions that involve all actors in the PFB process are highly promising to facilitate PA practices.

To conclude, we are convinced that this dissertation provides fruitful avenues for future research in order to further tailor the PFB process in function of students' learning. As such, it is hoped that researchers, instructional designers and educational practitioners are inspired by the studies in this dissertation and encouraged to implement well-considered collaboration scripts during PA practices. In this respect, PA practices should take into account a healthy mix of instructional support for all actors that involved in the PFB process, by integrating multiple practice occasions and triggering PFB dialogue.

References

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38*, 92-l13.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instruction effect of feedback in test-like events. *Review of Educational Research, 6*, 218-238.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151–167.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*, 474-482.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in higher education, 31*, 219-233.

Carless, D. (2015). *Excellence in University Assessment: Learning from Award-winning Practice*. London: Routledge.

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328–338.

Coll, C., Rochera, M.J., & De Gispert, I. (2014). Supporting online collaborative learning in small groups: teacher feedback on learning content, academic task and social participation. *Computers & Education, 75*, 53-64.

Creswell, J. W., & Garrett, A. L. (2008). The" movement" of mixed methods research and the role of educators. *South African Journal of Education, 28*, 321-333.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen: Open University of the Netherlands.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research in computer-supported collaborative learning: From design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3–19). Netherlands: Springer.

Dippold, D. (2009). Peer feedback through blogs: Student and teacher perceptions in an advanced German class. *ReCALL, 21*, 18-36.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83*, 70-120.

Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 128–143). New York: Routledge. G.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational psychologist, 48*, 56-66.

Flavell, J.H. (1976). Metacognitive aspects of problem solving. *The Nature of Intelligence, 12,* 231–235.

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.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315-325.

Hattie, J. & Gan, M. (2011). Instruction based on feedback. In Mayer, R. E. & Alexander, P. (eds). *Handbook of research on learning and instruction* (pp. 249-271). NewYork: Routledge. Taylor and Francis Group.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of educational research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152.

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

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., Hakkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2,* 211-224.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344–348.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*, 279-308.

Laurillard, D. (2002). Rethinking teaching for the knowledge society. *EDUCAUSE review, 37*, 16-24.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & quantity, 43*, 265-275.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*, 525-536.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* NE: Center for Instructional Innovation, University of Nebraska-Lincoln.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125-143). Mahwah, NJ: Erlbaum.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass

higher education. *Assessment and Evaluation in Higher Education, 35*, 501–517.

Nicol, D.J. and Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice. In C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. Ross & B. Smith (Eds.) *Enhancing Student Learning Through Effective Formative Feedback* (pp. 3-14). York: The Higher Education Academy.

Paterson, L., & Goldstein, H. (1991). New statistical methods for analysing social structures: an introduction to multilevel models. *British educational research journal, 17*, 387-393.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17*, 465-477.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–44.

Sadler, D.R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*, 535–550.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). An assessment for learning system called ACED: Designing for learning effectiveness and accessibility. *ETS Research Report Series, 2007*, i-45.

Sluijsmans, D. (2002). *Student involvement in assessment: the training of peer assessment skills*, unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages.* Paper presented at the meeting of EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

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

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*, 20–27.

Van der Pol, J., van den Berg, B. a. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*, 1804-1817.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Berg, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher- order peer feedback in EFL. *Learning and Instruction, 20*, 316–327.

Veenman, M.V.J., Kerseboom, L. & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availablity versus production deficiencies. *Anxiety, Stress, and Coping, 13*, 391–412.

Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education, 40*, 232-247.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.

Nederlandstalige samenvatting
Summary in Dutch

Nederlandstalige samenvatting

De impact van het structureren van peer feedback in een wiki-gebaseerde CSCL omgeving op de prestatie en de inhoud van de feedback.

Theoretische achtergrond

In het hedendaagse onderwijs wordt het geven en ontvangen van feedback alsmaar meer beschouwd als een interessante onderwijskundige activiteit die voordelen biedt voor zowel het leerproces (eg. Falchikov, 1995; Topping, 1998), als de prestaties van studenten (eg. Falchikov, 2003; Van Zundert, Sluijsmans, & Van Merriënboer, 2010). Wanneer studenten gevraagd worden om peer feedback (PFB) aan elkaar te geven, verwachten we van hen dat ze op een actieve manier omgaan met de opgelegde evaluatiecriteria van een opdracht om zo tot een objectief oordeel te komen over de kwaliteit van het werk van anderen. Anderzijds wordt er ook verwacht dat ze suggesties formuleren om het werk te verbeteren (Nicol, Thomson, & Breslin, 2014; Strijbos & Sluijsmans, 2010). In het kader van formatieve assessment moeten lerenden daarom de kans krijgen om met deze specifieke vorm van feedback aan de slag te gaan zodat ze enerzijds de kans krijgen hun prestatie te herwerken en anderzijds hun eigen leerproces kunnen bijsturen. Het geven en ontvangen van PFB wordt dan ook gezien als het leeraspect van peer assessment (PA) (Liu & Carless (2006), waarbij de feedback zelf wordt bestempeld als de essentiële factor voor 'verbetering' (Carless, 2015). Hierdoor zal de focus van dit proefschrift liggen op het geven en ontvangen van PFB, wat twee essentiële onderdelen van het PFB proces zijn tijdens PA praktijken.

Alsmaar meer onderzoek beklemtoont dat PA bijdraagt aan de leerwinst van zowel de beoordelaar als de beoordeelde, doordat PA praktijken niet enkel waardevol zijn als een assessment tool (Cheng & Warren, 1997), maar ook als een leermiddel op zichzelf (Topping, 1998). Er blijven echter nog vele vragen onbeantwoord over de wijze waarop deze vernieuwende assessmentvormen precies in de lespraktijk geïmplementeerd zouden moeten worden om het leerproces van studenten een 'boost' te geven (Sadler, 2010). In de literatuur wordt beklemtoond dat er verder onderzoek nodig is om te onderzoeken hoe lerenden tijdens PA praktijken het best ondersteund kunnen worden in het PFB proces om PFB van een hoge kwaliteit te garanderen (Hovardas, Tsivitanidou, & Zacharia, 2014). We kunnen er immers niet zomaar van uitgaan dat peers bekwaam genoeg zijn om feedback van een aanvaardbare kwaliteit te produceren (Kaufman & Schunn, 2010). Uit voorgaand onderzoek is gebleken dat

vooral de inhoud van de feedback in grote mate de effectiviteit van een PFB bericht bepaalt (Cho & MacArthur, 2010). Volgens het feedbackmodel van Hattie & Timperley (2007) is feedback het meest effectief wanneer het een antwoord biedt op drie belangrijke vragen: (1) Wat zijn de doelen? - 'Waar ga ik heen?' (feed up), (2) Welke vooruitgang richting het doel heb ik al geboekt? - 'Hoe ga ik?' (feedback), en (3) Welke activiteiten moet ik ondernemen voor een betere vooruitgang? - 'Naar waar ga ik nadien?' (feed forward). Om de inhoud meer gedetailleerd te onderzoeken, is er nood aan een efficiënt instrument om de werkelijke compositie van deze PFB onder de loep te nemen. Met dit proefschrift proberen we hierop in te spelen door het ontwikkelen en implementeren van een inhoudsanalyseschema voor het onderzoeken van de inhoud van PFB berichten, die lerenden aan mekaar geven tijdens schrijfopdrachten in een online leeromgeving in het hoger onderwijs.

In een cultuur van 'Assessment for Learning' moet het hedendaagse onderwijs dan ook ruimte voorzien voor deze vernieuwende, maar reeds algemeen aanvaarde assessmentvorm zoals PA (Evans, 2013; Nicol & Macfarlane-Dick, 2006). Uit de literatuur blijkt dat er geen eenduidige aanpak voor handen is om het PFB process te optimaliseren voor PA praktijken wanneer er PFB gegeven of ontvangen moet worden. (Strijbos & Sluijsmans, 2010). In de literatuur wordt aangeraden om te werken met 'collaboration scripts' of 'scripting' (Kollar & Fischer, 2010). Collaboration scripts geven meer details omtrent de roltoekenning bij groepsactiviteiten om studenten succesvol te laten samenwerken (Kollar, Fischer, & Hesse, 2006). Op deze manier wordt er voor de betrokken actoren op socio-cognitief vlak structuur voorzien door het specifiëren, plannen en delegeren van rollen en activiteiten (e.g. Fischer, Kollar, Stegmann, & Wecker, 2013). Dit gezegd zijnde, blijven er nog veel vragen onbeantwoord over hoe zulke collaboration scripts nu precies geïmplementeerd moet worden in de dagdagelijkse lespraktijk (Strijbos & Sluijsmans, 2010). In dit proefschrift proberen we hierop in te spelen door de rollen van de beoordelaar en de beoordeelde te structureren in het PFB proces om zo een dialoog te creëren. Hierbij focussen we hoofdzakelijk op de impact van het structureren van het PFB proces op de kwaliteit van de schrijfprestatie en PFB berichten tijdens PA praktijken in het eerste jaar van het hoger onderwijs.

Onderzoeksdoelen en –design

Het hoofddoel van dit proefschrift is om meer inzicht te krijgen in de manier waarop PA praktijken geoptimaliseerd kunnen worden en meer in het bijzonder hoe het structureren van de rol van de beoordeelde en de beoordelaar in het PFB proces een impact kan hebben op de kwaliteit van de prestatie (RO1) en de peer feedback (RO2) van lerenden. Hierop verder bouwend is het derde onderzoeksdoel: de ontwikkeling en implementatie van een inhoudsanalyseschema om de werkelijke inhoud van de PFB berichten gedetailleerder te analyseren, wat dan ook gebruikt wordt om de impact van de aangeboden structuur voor de

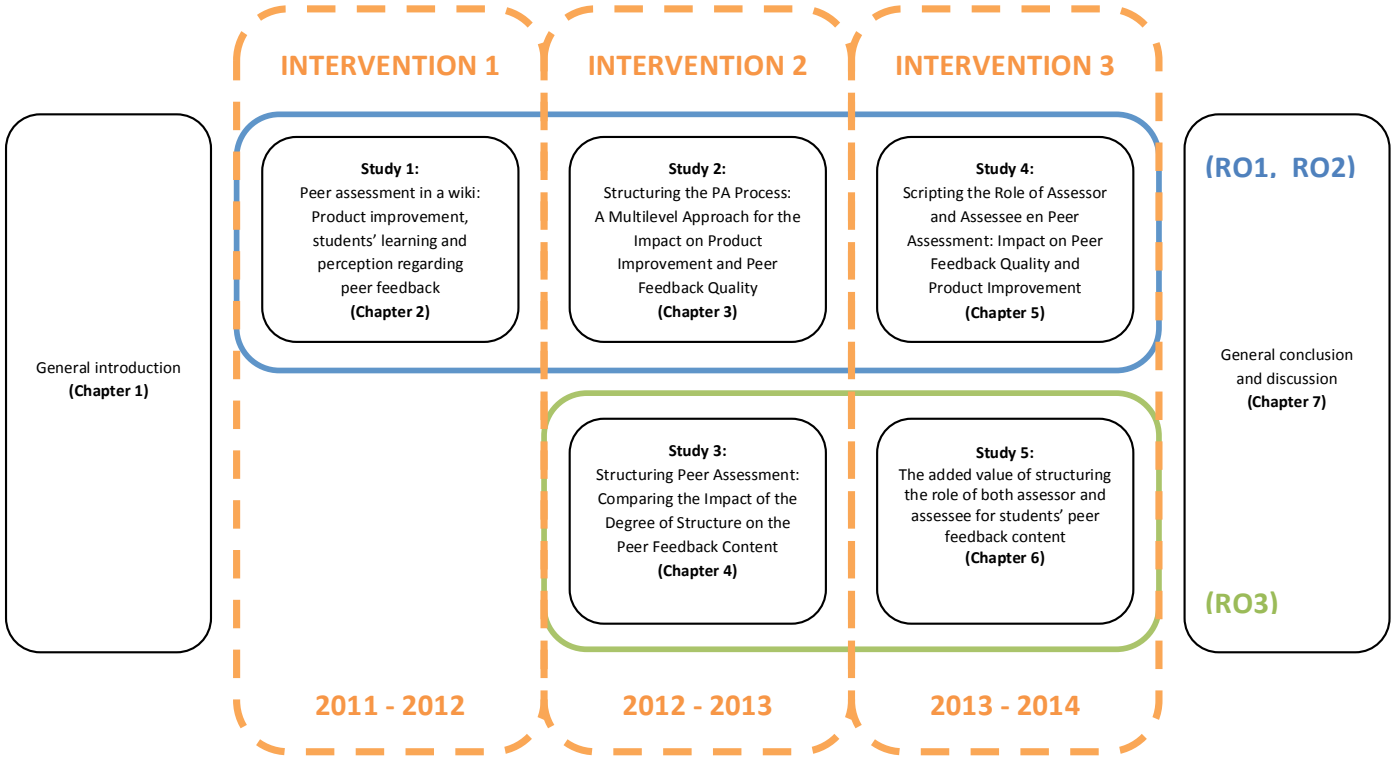
beoordeelde en de beoordelaar in het PFB proces te bestuderen m.b.t. de specifieke feedbackinhoud (RO3). Deze drie algemene onderzoeksdoelen vormen een leidraad doorheen de verschillende studies van dit proefschrift. In hoofdstuk 1 wordt eerst de theoretische achtergrond geschetst en nadien wordt een overzicht gegeven van de verschillende studies om uit te klaren hoe de verschillende hoofdstukken aan elkaar gerelateerd zijn. Zoals geïllustreerd in onderstaande Figuur 1, worden de resultaten van dit proefschrift beschreven in 5 studies (studie 1 t.e.m. 5, respectievelijk terug te vinden in hoofdstuk 2 t.e.m. 6), waarin de resultaten van 3 interventies met een quasi-experimenteel design besproken worden. Bij de tweede en de derde interventie worden telkens twee studies uitgevoerd, waarbij de eerste focust op OZD1 en OZD2, terwijl de tweede studie focust op OZD3. Tenslotte wordt er in de algemene discussie in hoofdstuk 7 beschreven hoe de bevindingen van de verschillende studies gerelateerd zijn aan de vooropgestelde onderzoeksdoelen.

Onderzoeksdoel 1 (OZD1): Het onderzoeken van de impact van een verschillend niveau van structurering van het PFB proces op de prestatie van de lerende.

Onderzoeksdoel 2 (OZD2): Het onderzoeken van de impact van een verschillend niveau van structurering in het PFB proces op de PFB kwaliteit van de lerende, gemeten door de Feedback Quality Index (Prins, Sluijsmans & Kirschner, 2006).

Onderzoeksdoel 3 (OZD3): Het ontwikkelen van een inhoudsanalyseschema dat gebruikt kan worden om de specifieke aard van inhoud van PFB berichten te onderzoeken, en het implementeren van dit inhoudsanalyseschema om de impact van een verschillend niveau van structurering van het PFB proces op de inhoud van de PFB berichten te bestuderen.

We willen benadrukken dat we de kwaliteit van de PFB dus op 2 manieren benaderen in dit proefschrift. Enerzijds bepalen we een kwaliteitsscore in functie van onderzoeksdoel 2, waarbij de kwaliteit van de feedback gemeten wordt door de Feedback Quality Index (Prins, Sluijsmans & Kirschner, 2006). Anderzijds bestuderen we de specifieke inhoud van de PFB berichten in meer detail door middel van een ontworpen inhoudsanalyseschema in functie van onderzoeksdoel 3.



Figuur 1. Overzicht van de 3 verschillende interventies en 5 studies

Onderzoeksdoelen 1 en 2 worden behandeld in studie 1 (hoofdstuk 2), studie 2 (hoofdstuk 3) en studie 4 (hoofdstuk 5). Hierin worden verschillen in de productscores en kwaliteitsscores van de PFB berichten onderzocht wanneer een verschillend niveau van structurering wordt aangeboden in het PFB proces.

In hoofdstuk 2 werden de deelnemers door de eerste interventie opgedeeld in 2 condities: een gestructureerde groep (S-PFB) en een niet-gestructureerde groep (controlegroep), waarbij voor de gestructureerde groep 2 extra richtvragen ('Wat vind je goed?' en 'Wat zou je veranderen?') zijn opgenomen in de aangereikte PFB template. In dit hoofdstuk worden de verschillen bestudeerd betreffende (1) het leerproces van de lerende, gemeten door meerkeuzevragen bij de aanvang en de afronding van de volledige opdrachtperiode, (2) de productverbetering van de oorspronkelijke draft versie naar de finale versie en (3) de perceptie van de lerende tijdens het PA proces in een CSCL omgeving in het hoger onderwijs. Hiervoor worden de data van 179 studenten, die worden gegroepeerd in 38 groepen, verzameld en geanalyseerd via variantieanalyses en independent-samples t-tests.

In hoofdstuk 3 worden de resultaten geanalyseerd van de deelnemers die een PFB template ontvingen, dewelke vooreerst een verschillend niveau van structurering in functie van de tweede interventie bevatte, wat resulteerde in 3 condities: geen structuur (controlegroep), (2) een basisstructuur en (3) een uitgebreide structuur. In de basisstructuur worden wederom 2 extra richtvragen ('Wat vind je goed?' en 'Wat zou je veranderen?') opgenomen in de PFB template, terwijl de conditie met uitgebreide structuur een PFB template ontving waarbij de

criteria georganiseerd werden volgens de principes van feed up, feedback en feed forward (Hattie & Timperley, 2007). In dit hoofdstuk wordt de impact van een verschillend niveau van structuur in de PFB template op de productscores en kwaliteitsscores van de PFB bestudeerd. Gebaseerd op deze resultaten, rapporteert hoofdstuk 3 ook over de ontwikkeling van een geoptimaliseerde PFB template die gebruikt kan worden door de beoordelaar om op een gestructureerde manier feedback te formuleren. Hiervoor wordt over 3 meetmomenten de data m.b.t. de productscore en kwaliteitsscore van de PFB van 168 studenten, gegroepeerd 37 groepen, verzameld en geanalyseerd via multilevel analyses.

In hoofdstuk 5 worden zowel de beoordeelde als de beoordelaar doelbewust actiever betrokken in het PFB proces. Dit is het grote verschil met hoofdstuk 3, aangezien hier enkel gefocust wordt op het structureren van de rol van de beoordelaar. Bij de derde interventie wordt er van de beoordeelde verwacht zelf te vragen naar specifieke feedback (PFB request) en anderzijds verwacht men dat hij de ontvangen feedback zal beoordelen. Tegelijkertijd wordt van de beoordelaar verwacht dat hij zijn feedback voorbereidt (content checklist), alvorens zijn feedback te formuleren a.d.h.v. een template die voorzien wordt door de instructor. Doordat hoofdstuk 5 verder bouwt op de bevindingen van hoofdstuk 3, gebruiken alle condities dezelfde geoptimaliseerde en gestructureerde PFB template. Bij deze interventie worden de deelnemers opgedeeld in 4 condities: (1) een controleconditie, (2) een feedback request conditie, (3) een content checklistconditie, en (4) een combinatieconditie (feedback request + content checklist). In hoofdstuk 5 wordt de impact van een verschillend niveau van structurering voor de rol van beoordelaar en beoordeelde in het PFB proces op de productscores en kwaliteitsscores bestudeerd. Hiervoor wordt over 3 meetmomenten de data van 125 studenten, gegroepeerd in 27 groepen, verzameld en geanalyseerd via multilevel analyses.

In functie van het derde onderzoeksdoel, wordt de specifieke inhoud van de PFB berichten die lerenden aan elkaar geven in meer detail bestudeerd. Hiervoor wordt er enerzijds in hoofdstuk 4 een nieuw inhoudsanalyseschema ontwikkeld dat gebruikt kan worden om de specifieke inhoud van PFB berichten te onderzoeken. Dit schema wordt nadien nog verder uitgebreid in hoofdstuk 6. Anderzijds wordt dit inhoudsanalyseschema geïmplementeerd om de impact van een verschillend niveau van structurering in het PFB proces te bestuderen met betrekking tot de specifieke inhoud van de PFB berichten tijdens schrijfopdrachten in een CSCL omgeving in het hoger onderwijs.

In hoofdstuk 4 worden inhoudelijke verschillen in de PFB berichten bestudeerd wanneer de beoordeelde een PFB template met een verschillende niveau van structurering gebruikt in het PFB proces (interventie 2). Om deze specifieke inhoud in meer detail te bekijken, rapporteert hoofdstuk 4 over de proportie van de verschillende PFB inhoudscategorieën: (1) peer feedbackstijl, (2) verificatietype, (3) verificatiefocus, (4) elaboratietype, en (5) elaboratiefocus Hiervoor worden 123 feedbackberichten van 41 studenten uit 9 groepen gesegmenteerd en gecodeerd. Dit resulteert in een database van 4717 feedbacksegmenten die geanalyseerd worden via variantieanalyses.

In hoofdstuk 6 wordt gekeken naar inhoudelijke verschillen wanneer zowel de beoordeelde (PFB request) als de beoordelaar (content checklist) bijkomende structuurparameters krijgen, bovenop de hulp van een gestructureerde PFB template (interventie 3). Hoofdstuk 6 focust op dezelfde PFB inhoudscategorieën als hoofdstuk 4, maar bekijkt daarnaast ook of de lerenden werkelijk akkoord gaan met de ontvangen PFB en of ze nadien de overeenkomstige aanpassingen maken. Hiervoor worden de categorieën (6) evaluatie-akkoord en (7) evaluatie-implementatie toegevoegd aan het inhoudsanalyseschema. Om dit te kunnen verwezenlijken worden 237 feedbackberichten van 79 studenten uit 16 groepen gesegmenteerd en gecodeerd. Dit resulteert in een database van 8440 feedbacksegmenten die geanalyseerd worden via multilevel analyses.

Overzicht van de resultaten

Onderzoeksdoel 1 (OZD1):

Het onderzoeken van de impact van een verschillend niveau van structurering in het PFB proces op de prestatie van de lerende.

Om een antwoord te kunnen formuleren op het eerste onderzoeksdoel wordt de impact op de schrijfprestatie van de lerende bestudeerd wanneer een verschillend niveau van structurering in het PFB proces geïmplementeerd wordt. Bij studie 1 wordt er van de studenten verwacht dat ze vroegere voorbeeldexamenvragen oplossen over drie belangrijke onderwerpen van de cursus: 'Behaviorism', 'Cognitivism' and 'Constructivism'. In de overige studies moeten studenten gedurende drie opeenvolgende opdrachten deelnemen aan schrijfopdrachten met de bedoeling om academische samenvattingen te schrijven die gebaseerd zijn op wetenschappelijke artikels. Bij iedere studie wordt de draft versie van de schrijfprestatie geëvalueerd door een peer om nadien deze PFB te kunnen gebruiken om de finale versie van de schrijfprestatie te verbeteren. Hieronder bespreken we de bevindingen van studie 1, 2 en 4.

In studie 1 (hoofdstuk 2) wordt de toegevoegde waarde van structuur tijdens het PFB proces bestudeerd wanneer de beoordelaar een PFB template ontvangt met 2 extra richtvragen ('Wat vind je goed?' en 'Wat zou je veranderen?'). Dit resulteert in 2 condities: (1) een gestructureerde groep (S-PFB) en (2) een niet-gestructureerde groep (controlegroep). Aangezien studenten verplicht worden om zowel een kladversie (draft) als een finale versie van hun schrijfprestatie in te dienen, tonen de resultaten aan dat de productscores significant verbeteren voor alle studenten wanneer ze de gelegenheid krijgen om hun draftversie te herwerken alvorens hun finale versie in te dienen. Het toevoegen van structuur aan een PFB template blijkt hierop echter

geen bijkomende invloed te hebben. Aan de andere kant illustreren de resultaten van studie 1 wel dat studenten, die PFB geven en ontvangen m.b.v. een gestructureerde PFB template, zich kritischer opstellen t.a.v. de PFB en dat ze deze ontvangen feedback ook als diepgaander en meer gedetailleerd beschouwen.

Studie 2 (hoofdstuk 3) bekijkt wat de impact is op de schrijfprestaties van de lerende, wanneer er een verschillend niveau van structurering aangeboden wordt in een PFB template in het PFB proces. In deze studie zijn er 3 condities: geen structuur (controlegroep), (2) een basisstructuur met 2 richtvragen ('Wat vind je goed?' en 'Wat zou je veranderen?') en (3) een uitgebreide structuur waarbij de evaluatiecriteria gegroepeerd en herhaald worden volgens de feedbackprincipes van Hattie & Timperley (2007) (Feed up - 'Where am I going?', Feedback - 'How am I going?' en Feed forward - 'Where to next?'). De resultaten tonen aan dat de kwaliteit van de finale versie van schrijfprestaties significant verbetert voor alle lerenden. Met betrekking tot het verschillende niveau van structurering in de PFB template, wordt enkel op het laatste meetmoment een significant verschil vastgesteld tussen de verschillende condities. In vergelijking met lerenden die helemaal geen extra structuur ontvangen in het PFB proces, hebben lerenden met zowel een basis als uitgebreide structuur significant hogere productscores na meerdere oefenmomenten. Verder worden geen significante verschillen meer gevonden tussen de condities. Op deze manier illustreert studie 2 dat het aanbieden van extra structuur in een PFB template een meerwaarde kan betekenen voor de schrijfprestatie van de lerende.

Studie 4 (hoofdstuk 5) bestudeert op welke manier een interventie gunstig kan zijn voor de kwaliteit van het geschrevene, wanneer zowel de beoordeelde als beoordelaar actiever betrokken worden in het PFB proces. Dit door middel van verdere instructies die deel uitmaken van interventie 3. Hierdoor moet de beoordeelde vragen naar specifieke feedback en deze nadien ook evalueren. Daarnaast moet de beoordelaar zijn feedback voorbereiden vooraleer deze te formuleren in een aangereikte en geoptimaliseerde PFB template. Deze studie heeft een 2x2 factorial design met 4 condities: (1) een controleconditie, (2) een feedback request conditie, (3) een content checklistconditie, en (4) een combinatieconditie (feedback request + content checklist). De resultaten van studie 3 tonen aan dat alle lerenden na verloop van tijd erin slagen om hun schrijfprestatie significant te verbeteren. Wanneer we meer in detail kijken naar de verschillende meetmomenten, wordt duidelijk dat de kwaliteit van de schrijfprestatie significant verbetert van tijd 1 naar tijd 2 voor beide condities die een PFB request gebruiken in het PFB proces. Aangezien de productscores ook voor de controleconditie significant stijgen, kan deze bevinding de werkelijke behoefte aan extra structuur of ondersteuning in het PFB proces in functie van de schrijfprestatie in vraag stellen.

Belangrijk om op te merken is dat het PFB proces in zekere mate een stapsgewijze procedure oplegt voor alle lerenden, aangezien alle condities dezelfde gestructureerde PFB template gebruiken gedurende de gehele opdrachtperiode.

Onderzoeksdoel 2 (OZD2):

Het onderzoeken van de impact van een verschillend niveau van structurering in het PFB proces op de PFB kwaliteit van de lerende, gemeten door de Feedback Quality Index.

In het kader van het tweede onderzoeksdoel wordt de impact op de kwaliteit van de PFB berichten van de lerende bestudeerd wanneer een verschillend niveau van structurering van het PFB proces geïmplementeerd wordt. Hieronder bespreken de bevindingen van studie 2 en 4. In functie van OZD2 benaderen we de kwaliteit van de PFB berichten op basis van een kwaliteitsscore die gemeten wordt door de Feedback Quality Index (Prins, Sluijsmans & Kirschner, 2006).

In studie 2 (hoofdstuk 3) wordt gekeken wat de impact op de kwaliteit van de PFB is, wanneer er een PFB template wordt aangereikt met (1) geen structuur (controlegroep), (2) een basisstructuur en (3) een uitgebreide structuur. Naarmate de tijd vordert, illustreren de resultaten voor alle lerenden een globale significante stijging van de kwaliteit van de PFB. Met betrekking tot het verschillende niveau van structurering in de PFB template, tonen de resultaten van studie 2 aan dat studenten, die een uitgebreide structuur ontvangen in hun PFB template, significant hogere PFB kwaliteitsscores hebben in vergelijking met studenten die geen extra structuur (controlegroep) of enkel 2 richtvragen (basisstructuur) krijgen in hun PFB template. Het aanbieden van een PFB template met een uitgebreide structuur waarbij de evaluatiecriteria gegroepeerd en herhaald worden volgens de feedbackprincipes van Hattie & Timperley (2007) (Feed up - 'Where am I going?', Feedback - 'How am I going?' en Feed forward - 'Where to next?'), blijkt dus een effectieve manier om de kwaliteitsscores van de PFB te laten stijgen.

Studie 3 (hoofdstuk 5) onderzoekt de impact op de kwaliteit van de PFB, wanneer lerenden actiever betrokken worden in het PFB proces doordat ze als beoordeelde en beoordelaar verdere instructies krijgen. De beoordeelde wordt opgedragen om te vragen naar specifieke feedback m.b.t. bepaalde criteria, terwijl de beoordelaar eerst moet deelnemen aan een voorbereidende taak alvorens feedback geformuleerd kan worden op een gestructureerde PFB template, die gelijk is voor alle lerenden. Dit resulteert in 4 condities: (1) een controleconditie, (2) een feedback request conditie, (3) een content checklistconditie, en (4) een combinatieconditie (feedback request + content checklist). De resultaten van studie 5 tonen aan dat naarmate de tijd vordert alle lerenden PFB-berichten van een hogere kwaliteit formuleren. Vanaf meetmoment 2, worden bepaalde condities gestructureerd zodat de beoordeelde moet vragen naar specifieke feedback (PFB request) en de beoordelaar een voorbereidingsopdracht (content checklist) moet maken alvorens feedback op de PFB template te formuleren. Wanneer we in meer detail naar de verschillende meetmomenten kijken, zien we dat bij meetmoment 1 er reeds significante verschillen zijn tussen de condities voor de start van de interventie. Meer bepaald tonen de resultaten significant hogere PFB kwaliteitsscores op meetmoment 1 in beide

condities waar een content checklist deel uitmaakt van het PFB proces. Dit kan op zijn beurt ook een verklaring zijn voor de niet-significante, stelselmatige stijging voor deze condities. Hierdoor kunnen de bevindingen omtrent de impact van een content checklist als weinig doorslaggevend bestempeld worden. Het vragen naar specifieke PFB blijkt dan weer wel een nuttige manier te zijn om de kwaliteit van de PFB te laten stijgen, en dit vooral in de beginfase van de opdrachtperiode. Beide condities, waarbij de beoordeelde moet vragen naar specifieke feedback m.b.v. een PFB request, vertonen een significante kwaliteitsstijging van meetmoment 1 naar 2. De resultaten illustreren dat enkel de controleconditie significant stijgt van meetmoment 2 naar 3. Aangezien de significante verschillen in PFB kwaliteitsscores van meetmoment 1 allemaal zijn verdwenen op meetmoment 2 en 3, suggereren deze bevindingen dat alle lerenden bekwamere beoordelaars worden wanneer ze meerdere gelegenheden krijgen om te oefenen.

Onderzoeksdoel 3:

Het ontwikkelen van een inhoudsanalyseschema dat gebruikt kan worden om de specifieke aard van inhoud van PFB berichten te onderzoeken, en het implementeren van dit inhoudsanalyseschema om de impact van een verschillend niveau van structurering van het PFB proces op de inhoud van de PFB berichten te bestuderen.

Bij de start van dit onderzoek, zijn er in de literatuur nog geen relevante inhoudsanalyseschema's beschikbaar waarmee de inhoud van PFB berichten in deze specifieke context in detail bestudeerd kunnen worden. Om het laatste onderzoeksdoel te bereiken, zijn we dus voornamelijk geïnteresseerd in de specifieke inhoud van de PFB-berichten die lerenden met mekaar uitwisselen. Geïnspireerd door een recent ontwikkeld codeerschema (Strijbos, Van Goozen, & Prins, 2012), dat op zijn beurt gebaseerd is op het algemeen aanvaard theoretisch kader rond feedback (Narciss, 2008), richt het derde onderzoeksdoel zich op de ontwikkeling van een inhoudsanalyseschema om de specifieke inhoud van de PFB berichten van lerenden te analyseren. De focus ligt hierbij vooral op PFB berichten die gebruikt worden tijdens schrijfopdrachten in een CSCL-omgeving in het hoger onderwijs.

In studie 3 (hoofdstuk 4) wordt de impact op de kwaliteit van de PFB-inhoud verder uitgediept, meer specifiek wanneer er een PFB template met een verschillende structureringsgraad geïmplementeerd wordt in het PA proces. Door het uitvoeren van kwantitatieve inhoudsanalyse en variatieanalyses, probeert hoofdstuk 4 meer licht te werpen op mogelijke verschillen tussen de 3 condities: (1) geen structuur (controlegroep), (2) een basis structuur en (3) een uitgebreide structuur. Daarom ligt de focus op de volgende PFB inhoudscategorieën: (1) peer feedbackstijl, (2) verificatietype, (3) verificatiefocus, (4) elaboratietype, en (5) elaboratiefocus, wanneer een PFB template met een verschillende niveau

van structurering is geïmplementeerd in het PFB proces. De resultaten van hoofdstuk 4 illustreren dat globaal gezien alle lerenden een gebalanceerde proportie van verificaties en elaboraties voorzien in hun PFB-berichten. Het toevoegen van enkele richtvragen als basisstructuur in de PFB template blijkt de proportie elaboraties significant te verhogen in vergelijking met studenten die voor de rest geen verdere structuur ontvangen. Met betrekking tot het feedback type, tonen de resultaten dat lerenden grotendeels positieve verificaties formuleren. Daarnaast geven ze ook een gebalanceerde proportie van suggestieve en informatieve elaboraties. Bijkomend tonen de resultaten aan dat de PFB-berichten zich grotendeels focussen op bepaalde criteria. Wanneer lerenden een PFB template met een uitgebreide structuur gebruiken, leren de resultaten ons dat ze meer geneigd zijn om feedback te formuleren op het criteriumniveau. Wanneer lerenden een PFB template zonder enige structuur aangereikt krijgen, blijken ze sneller geneigd om meer algemene feedback te geven over de taak in zijn geheel of eerder op taalaspecten, waarin ze hoofdzakelijk bevestigen of de prestatie overeenkomt met het verwachte resultaat of niet. Hierdoor is het dan ook aan te raden voor instructiegevers om wanneer er een PFB template wordt voorzien, een lijst van ondubbelzinnige criteria mee te geven, waardoor beoordelaars gestructureerd worden in het PFB proces om deze lijst te toetsen aan de feedback principes van feed up, feedback en feed forward (Hattie & Timperley, 2007).

Als laatste onderdeel van dit proefschrift, wordt in studie 5 (hoofdstuk 6) de impact van een interventie bestudeerd, waarbij zowel de beoordeelde (vragen en evalueren van feedback) als de beoordelaar (voorbereiden en formuleren van feedback) bijkomende instructies ontvangen, op de specifieke PFB inhoud tijdens meerdere PFB cycli in functie van schrijfopdrachten in het eerste jaar van het hoger onderwijs. Zoals voordien aangegeven, resulteert dit in 4 condities: (1) een controleconditie, (2) een feedback request conditie, (3) een content checklistconditie, en (4) een combinatieconditie (feedback request + content checklist). In hoofdstuk 6 ligt de focus op de volgende PFB inhoudscategorieën: (1) peer feedbackstijl, (2) verificatietype, (3) verificatiefocus, (4) elaboratietype, (5) elaboratiefocus, (6) evaluatie-akkoord en (7) evaluatie-implementatie. De resultaten tonen aan dat de aanvullende structuur voor de beoordeelde en beoordelaar geen diepgaand effect heeft op de inhoud van de PFB berichten. Het enige significante verschil tussen de condities is dat het vragen naar feedback (PFB request) en/of het voorbereiden van feedback (content checklist) een nuttige manier blijkt te zijn om de informatieve component in PFB berichten te verhogen, vooral wanneer lerenden meerdere gelegenheden krijgen om te oefenen. Bij deze informatieve elaboraties gaan lerenden uitgebreider in op de positieve of negatieve verificaties die ze voordien m.b.t. een bepaald criterium gemaakt hebben. Daarnaast worden geen significante verschillen betreffende de specifieke PFB inhoud opgemerkt. De resultaten illustreren dat globaal gezien alle lerenden een redelijk gebalanceerde proportie van verificaties en elaboraties voorzien in hun PFB-berichten; hetgeen vergelijkbaar is met de resultaten van hoofdstuk 4. Met betrekking tot het feedback type, tonen de resultaten aan dat lerenden grotendeels positieve verificaties formuleren die zich focussen op bepaalde criteria. Daarnaast formuleren ze ook een gebalanceerde hoeveelheid informatieve en suggestieve elaboraties, die wederom voornamelijk betrekking hebben op bepaalde criteria.

Overigens onthullen de resultaten van hoofdstuk 6 ook dat alle condities naarmate de tijd vordert minder negatieve verificaties verwoorden in de PFB berichten. De kwaliteit van de schrijfprestaties stijgt stelselmatig significant hoger voor alle lerenden, nadat ze meerdere gelegenheden hebben gekregen om aan een gelijkaardige opdracht te werken. Aangezien het in studie 3 ook de bedoeling is om de 'feedback loop' te sluiten (Boud, 2000), wordt er tevens gekeken of lerenden over het algemeen akkoord gaan met de feedback die ze ontvangen, en of de feedback werkelijk gebruikt wordt bij de revisie van de draftversie. De resultaten geven aan dat lerenden grotendeels akkoord gaan met de ontvangen feedback en dat ze bijna de helft van deze PFB berichten ook daadwerkelijk implementeren in hun finale versie.

Algemene conclusie

Na een uitgebreide literatuurstudie blijkt duidelijk dat er nood is aan meer onderzoek rond het optimaliseren van het PFB proces in functie van het leerproces van de lerende (e.g., Hovardas, Tsivitanidou, & Zacharia, 2014). Om bij te dragen tot dit leerproces, zijn we ervan overtuigd dat 'collaboration scripts' voor samenwerkend leren en assessmentactiviteiten weldoordacht moeten zijn. Recent nog werd het belang van een feedbackdialoog tussen de betrokken actoren benadrukt in de literatuur (Carless, 2015). Spijtig genoeg wordt de rol van de beoordeelde in het merendeel van de interventiestudies nog steeds over het hoofd gezien, waardoor er van een echte PFB dialoog in twee richtingen geen sprake is. In dit opzicht beperken de verschillende interventiestudies van dit proefschrift zich niet enkel tot het structureren van de rol van de beoordelaar om kwalitatieve PFB te bewaken, maar richt de laatste interventie zich ook op hoe de beoordeelde meer actief betrokken kan worden in dit PFB proces. Ook al beseffen we goed dat de vijf empirische studies die in dit proefschrift gepresenteerd worden enkel maar het topje van een nog veel grotere ijsberg laten zien, toch zijn we ervan overtuigd dat dit proefschrift een constructieve aanzet kan zijn voor toekomstig onderzoek naar instructie gevende interventies om het PFB proces te optimaliseren voor zowel de beoordelaar als de beoordeelde.

Bij iedere studie van dit proefschrift wordt aangetoond dat de kwaliteit van zowel de schrijfprestatie als de PFB berichten voor alle lerenden significant stijgt naarmate de tijd vordert. In de verschillende studies wordt duidelijk dat dit niet zozeer het gevolg is van de aangeboden structuur in het PFB proces, maar vooral van het aantal oefengelegenheden die de lerende krijgt bij gelijkaardige schrijfopdrachten en assessmentactiviteiten. Anderzijds toont studie 1 wel aan dat enkele richtvragen in een PFB template ervoor kunnen zorgen dat lerenden zich kritischer durven opstellen ten aanzien van de ontvangen de feedback van de beoordelaar, maar ook dat ze deze feedback als waardevollere en gedetailleerdere informatie gaan beschouwen. Tevens wordt bij alle lerenden een significante stijging van draft naar finale versie waargenomen. Daarenboven illustreert studie 2 dat de aanwezigheid van een basis of uitgebreide structuur in de PFB template, door het aanbieden van respectievelijk een aantal richtvragen of een criterialijst volgens de principes van feed up, feedback en feed forward, een meerwaarde kan betekenen en dus zorgt voor een toenemende kwaliteit van zowel de schrijfprestaties als de PFB. Wanneer beide actoren betrokken worden om de PFB dialoog te stimuleren, tonen de resultaten van studie 4 aan dat het nuttig is om de beoordeelde te laten formuleren op welke specifieke criteria PFB gewenst wordt. De integratie van een 'PFB request' zorgt voor een significante stijging van de PFB kwaliteit, en dit vooral bij de start van de opdrachtperiode.

Wanneer we in meer detail naar de werkelijke compositie van de PFB inhoud kijken, vertonen de resultaten van studie 3 en 5 een aantal gelijkenissen. Tevens wordt in deze hoofdstukken aangetoond dat de aanvullende structuur in het PFB proces geen bepalend effect heeft op de PFB inhoud. Het wordt in beide studies duidelijk dat studenten PFB berichten

formuleren met een gebalanceerde verhouding van verificaties en elaboraties. Effectieve feedback moet dan ook bestaan uit zowel verificaties als elaboraties (Narciss, 2008). Dit geeft aan dat lerenden zich over het algemeen niet beperken tot het aangeven of er wel of niet aan een bepaald criterium is voldaan, waarbij ze voornamelijk geneigd zijn om positieve feedback te geven, maar dat ze daarnaast ook consistent aangegeven wat de reden hiervoor is of wat er gedaan kan worden om de kwaliteit te verbeteren. De PFB berichten bestaan uit een vrij gebalanceerde verhouding van informatieve en suggestieve feedback. Vervolgens blijkt het toevoegen van richtvragen aan de PFB template een effectieve manier om de elaboratiecomponent in de PFB berichten te verhogen. Daarnaast wordt ook duidelijk dat een criterialijst volgens de principes van feed up, feedback en feed forward ervoor zou kunnen zorgen dat lerenden stelselmatiger feedback formuleren op criteriumniveau. Op basis van deze bevindingen, wordt in studie 4 en 5 aan alle lerenden eenzelfde gestructureerde PFB template aangereikt. Hieruit blijkt dat het aantal negatieve verificaties na verloop van tijd afneemt, wat zou verklaard kunnen worden door de stijgende kwaliteit van de schrijfprestatie.

Deze laatste studie illustreert ook dat het verder specifiëren van de rol van de beoordeelde en beoordelaar verder geen grote verschillen qua PFB inhoud oplevert, behalve dan dat de lerenden die meer structuur meekrijgen significant meer informatieve elaboraties verwerken in hun PFB berichten. Wanneer de beoordelaar aanvullende informatie voorziet om een specifiek oordeel te beargumenteren, delen wij de visie dat 'justifications' bevorderlijk zijn voor de schrijfprestatie van lerenden (e.g. Stern & Solomon, 2006) en beschouwen we dit als een essentieel element van formatieve assessment.

Dit proefschrift heeft ook enkele implicaties. Eerst en vooral wordt in dit proefschrift een inhoudsanalyse schema ontwikkeld om de specifieke feedbackinhoud die lerenden aan elkaar geven tijdens schrijfopdrachten en assessmentactiviteiten in een online leeromgeving in het hoger onderwijs, meer in detail te kunnen bestuderen. Dit inhoudsanalyseschema wordt geïmplementeerd in studie 3 en studie 5 en kan dus ingezet worden in toekomstige empirisch onderzoek dat de specifieke inhoud van PFB berichten onder de loep wil nemen. Ook hopen we dat onderzoekers verder zullen bouwen op het ontwikkelde inhoudsanalyseschema. Naast de ontwikkeling van dit schema illustreren de resultaten van studie 3 en 5 ook dat de PFB berichten in algemeen bestaan uit een gebalanceerd proportie van verificaties en elaboraties die voornamelijk focussen op criteriumniveau, waarvan de verificaties hoofdzakelijk positief zijn en de elaboraties evenwaardig bestaan uit informatieve en suggestieve verhelderingen. Studie 5 illustreert daarenboven dat lerenden meestal akkoord gaan met de ontvangen PFB, waarvan ze ongeveer de helft van de suggesties implementeren.

Daarnaast roepen de onderzoeksresultaten van de verschillende studies om een gebruiksvriendelijke en eenvoudige PFB template te implementeren bij PA praktijken. Op basis van de resultaten van studie 1, 2 en 3 wordt een geoptimaliseerde PFB template ontwikkeld die wordt geïmplementeerd in studie 4 en 5. Een dergelijke PFB template bestaat uit 4 componenten: een lijst met duidelijke criteria, een sectie om feedback te geven, een sectie om feed forward te formuleren en een sectie om de ontvangen feedback te evalueren. Deze template

kan gebruikt worden in toekomstige studies om de rol van de beoordelaar in het PFB proces te structureren. Hopelijk leidt dit proefschrift ook tot een belangrijke hernieuwde interesse bij onderzoekers naar gelijkaardige instructionele interventies in PA praktijken, die als doel hebben om het PFB proces van zowel de beoordelaar als de beoordeelde te optimaliseren in functie van het leerproces. Geïnspireerd door de interventies van dit proefschrift kunnen gelijkaardige toekomstige interventies opgezet worden om de impact op de kwaliteit van de prestatie en PFB te bestuderen. De resultaten van dit proefschrift benadrukken dat er vooral meerdere oefenmomenten en cycli voorzien moeten worden in het design van PA praktijken om de kwaliteit van de schrijfprestaties en PFB berichten significant te laten stijgen.

Dit proefschrift heeft echter ook beperkingen. Een eerste heeft betrekking op de steekproef waarop de multilevel en variantieanalyses worden uitgevoerd. Een uitbreiding van de steekproef is aan te raden om na te gaan of gelijkaardige resultaten en meer accurate schattingen en standaardfouten worden gevonden. Daarnaast kunnen deze interventies herhaald worden bij secundaire en master studenten, bij jongere en oudere studenten, met een grotere gender diversiteit en bij verschillende studentenpopulaties uit diverse cursussen in verschillende opleidingsprogramma's. Verder is de inhoud van de PFB request of de content checklist niet verder in detail onderzocht. Hiervoor zou het ontwikkelde inhoudsanalyseschema aangepast kunnen worden om de inhoud van de PFB request en content checklist te analyseren. Het zou tevens waardevol kunnen zijn om de relatie te tussen deze inhoud en de kwaliteit van de gegeven feedback te onderzoeken. Ook kunnen we dan nagaan hoe deze gegeven PFB uiteindelijk geëvalueerd wordt. Alhoewel het gebruik van kwalitatief onderzoek sterk wordt aangemoedigd in onderwijskundig onderzoek, beperkt dit proefschrift zich tot kwantitatief onderzoek. Verder onderzoek zou bijvoorbeeld via interviews of focusgroepen meer te weten kunnen komen over de perceptie van de studenten, wanneer ze ervaring opdoen in hun rol van beoordelaar en beoordeelde. Hierbij zou ook de impact op de schrijfprestatie onderzocht kunnen worden, wanneer studenten enkel deelnemen als beoordelaar of beoordeelde in het PFB proces.

Tot slot volgt een terugkoppeling naar het algemene doel van dit proefschrift namelijk inzicht verwerven in de manier waarop PA praktijken geoptimaliseerd kunnen worden, en meer in het bijzonder hoe het structureren van de rol van de beoordeelde en de beoordelaar in het PFB proces een impact kan hebben op de kwaliteit van de prestatie en de PFB van lerenden. De resultaten in dit proefschrift tonen aan dat een verschillend niveau van structurering in het PFB proces niet altijd een dermate grote rol invloed heeft op de kwaliteit van de schrijfprestaties en PFB berichten, maar dat het design van PA praktijken vooral meerdere oefenmomenten en cycli moet voorzien om de kwaliteit van de prestaties en PFB significant te laten stijgen. Door de ontwikkeling en implementatie van een inhoudsanalyseschema hebben we meer inzicht verworven m.b.t. de werkelijke inhoud van de PFB berichten die studenten aan elkaar geven. We verwijzen naar hoofdstuk 7 voor de algemene discussie van dit proefschrift, waarin de belangrijkste resultaten besproken worden rond drie terugkerende thema's: 'Structuring versus practicing the PFB process, 'Structuring the assessor's PFB template', en 'Triggering peer feedback dialogue'. Als laatste hopen we dat dit proefschrift interessante pistes biedt voor

toekomstig onderzoek en dat het een bron van inspiratie mag zijn voor onderwijskundige instellingen, departementen, onderzoekers, onderwijskundigen, instructionele ontwerpers en lesgevers.

Referenties

Carless, D. (2015). *Excellence in University Assessment: Learning from Award-winning Practice*. London: Routledge.

Cheng, W., & Warren, M. (1997). Having second thoughts: Students perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239.

Cho, K., & MacArthur, C. (2010) Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83*, 70-120.

Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 157-166). London: Kogan Page.

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching, 3*, 102–108.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of educational research, 77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133–152.

Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*, 387-406.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*, 159–185.

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

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*, 102-122.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*, 535-550.

Stern, L. A., & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing, 11*, 22-41.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269.

Strijbos, J. W., Van Goozen, B., & Prins, F. (2012, August). *Developing a coding scheme for analysing peer feedback messages*. Paper presented at the EARLI-SIG 1 Assessment and Evaluation Conference, Brussels, Belgium.

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

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270-279.

Academic output

Academic output

Journals (A1)

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315-325.

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education, 88*, 370-386.

De Wever, B., Hämäläinen, R., Voet, M., & Gielen, M. (2015). A wiki task for first-year university students: The effect of scripting students' collaboration. The Internet and Higher Education, 25, 37-44.

Gielen, M., De Wever, B., Voet, M. (Accepted). The added value of structuring the role of both assessor and assessee during multiple feedback cycles in a wiki environment for students' peer feedback content. Manuscript submitted for publication in the *Journal of Computer Assisted Learning.*

Conference papers (P1)

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. Procedia - Social and Behavioral Sciences, 69, 585-594.

Gielen, M., & Dewever, B. (2013). Structuring the PA process: impact on feedback quality. In N. Rummel, M. Kapur, N. Mitchell, & S. Puntambekar (Eds.), CSCL 2013 Conference Proceedings, 2, 255–256.

Conference contributions

Gielen, M. (2011). Learning (R)evolution – Learning of Generation Y. Presentation at the studiedag Vlaams Forum voor Onderwijsonderzoek (VFO), Ghent, Belgium, 17 November 2011.

Gielen, M. & De Wever, B. (2012). Peer-assessment in een Wiki: Het effect van gestructureerde peer feedback op het proces en product van groepswerk. Paper presented at the Onderwijs Research Dagen (ORD), Wageningen, The Netherlands, 20 – 22 June 2012.

Gielen, M. & De Wever, B. (2012). Peer-assessment in a Wiki: The effect of structured peer feedback on the process and product of group work. Paper presented at the International Conference on Education and Educational Psychology (ICEEPSY), Istanbul, Turkey, 10 – 12 October 2012.

Gielen, M. & De Wever, B. (2013). Peer-assessment in a Wiki: The effect of structured peer feedback on the process and product of group work. Poster presented at the Onderwijs Research Dagen (ORD), Brussel, Belgium, 29 – 31 May 2013.

Gielen, M. & De Wever, B. (2013). Structuring the PA process: impact on feedback quality. Poster presented at the Conference on Computer-supported Collaborative Learning (CSCL), Chicago, US, 15 – 16 June 2013.

Gielen, M. & De Wever, B. (2013). The added value of scaffolding the self and peer assessment process in a Wiki-based CSCL-environment in higher education. Paper presented at Doctoral Consortium at the Conference on Computer-supported Collaborative Learning (CSCL), Chicago, US, 15 – 16 June 2013.

Gielen, M. & De Wever, B. (2013). Structuring the Peer Assessment Process: The Impact on Product Improvement, Students' Perception and Feedback Quality. Paper presented at the EARLI conference, Munich, Germany, 27 – 31 August 2013.

Gielen, M. & De Wever, B. (2014). Structuring the Role of Assessor and Assessee in the Peer Assessment Process: The Impact on Product Improvement and Peer Feedback. Paper presented at the SIG1 conference on Professional Development in Assessment, Madrid, Spain, 27 – 29 June 2014.

Gielen, M. & De Wever, B. (2014). Structuring the Peer Assessment Process: A Multilevel Approach for the Impact on Product Improvement and Feedback Quality. Paper presented at the European Conference on Educational Research (ECER), Porto, Portugal, 2 – 5 September 2014.

Gielen, M. & De Wever, B. (2015). Does shifting the structuring degree of the peer assessment process influence students' peer feedback content? Paper presented at the conference for Assessment for Learning in Higher Education, Hong Kong, China, 14 – 15 May 2015.

Gielen, M. & De Wever, B. (2015). The impact of varying degrees of scripting on peer feedback quality and performance improvement. Paper presented at the Earli conference, Limassol, Cyprus, 25 – 29 August 2015.

Data storage facts sheets

% Data Storage Fact Sheet
% Study 1_Chapter 2
% Author: Mario Gielen
% Date: March, 24, 2016

1. Contact details
==========
1a. Main researcher

- name: Mario Gielen
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: mario.gielen@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Bram De Wever (Supervisor PhD Project)
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: bram.dewever@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Gielen, M., & De Wever, B. (2012). Peer assessment in a wiki: Product improvement, students' learning and perception regarding peer feedback. *Procedia-Social and Behavioral Sciences, 69*, 585-594.

* Which datasets in that publication does this sheet apply to?:
This sheet applies to the complete dataset of the study reported in Chapter 2 of the dissertation.

3. Information about the files that have been stored
==========
3a. Raw data

The raw data consist of individual pre-and post-test scores related to the six multiple-choice exam questions, scores on the wiki product, and data from the questionnaire at the end of the wiki-assignment on students' perception towards the PA process.

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:
* On which platform are the raw data stored?
 -[ ] researcher PC
 -[x] research group file server

-[x] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

3b. Other files
---
* Which other files have been stored?
-[ ] file(s) describing the transition from raw data to reported results. Specify:
-[x] file(s) containing processed data. Specify: All the gathered data was cleaned and aggregated for analysis in SPSS).
-[x] file(s) containing analyses. Specify: SPSS-generated output was stored (i.e. output of preliminary analyses as well as output of the main analyses regarding the research questions).
-[ ] files(s) containing information about informed consent
-[ ] a file specifying legal and ethical provisions
-[ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
-[ ] other files. Specify: ...

* On which platform are these other files stored?
-[x] individual PC
-[x] research group file server
-[ ] other:
* Who has direct access to these other files (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
- name:
- address:
- affiliation:
- e-mail:

% Data Storage Fact Sheet
% Study 2_Chapter 3
% Author: Mario Gielen
% Date: March, 24, 2016

1. Contact details
==========
1a. Main researcher

- name: Mario Gielen
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: mario.gielen@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Bram De Wever (Supervisor PhD Project)
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: bram.dewever@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449.

* Which datasets in that publication does this sheet apply to?:
This sheet applies to the complete dataset of the study reported in Chapter 3 of the dissertation.

3. Information about the files that have been stored
==========
3a. Raw data

The raw data consist of the peer feedback messages and of the draft and final version of the wiki product, which were measured with scoring rubrics.

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:
* On which platform are the raw data stored?
 -[ ] researcher PC
 -[x] research group file server

-[x] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

3b. Other files
---
* Which other files have been stored?
-[x] file(s) describing the transition from raw data to reported results. Specify: The rubrics to score the quality of the peer feedback messages and the wiki product.
-[x] file(s) containing processed data. Specify: The data was cleaned in SPSS, aggregated for analysis and restructured for multilevel analysis using MLwiN).
-[x] file(s) containing analyses. Specify: all MLwiN 2.23-generated model outputs (i.e. output of priliminary analyses as well as output of the main analyses regarding the research questions) were stored as .wsz files.
-[ ] files(s) containing information about informed consent
-[ ] a file specifying legal and ethical provisions
-[ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
-[ ] other files. Specify: ...

* On which platform are these other files stored?
-[x] individual PC
-[x] research group file server
-[ ] other:
* Who has direct access to these other files (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
- name:
- address:
- affiliation:
- e-mail:

% Data Storage Fact Sheet
% Study 3_Chapter 4
% Author: Mario Gielen
% Date: March, 24, 2016

1. Contact details
==========
1a. Main researcher

- name: Mario Gielen
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: mario.gielen@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Bram De Wever (Supervisor PhD Project)
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: bram.dewever@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315-325.

* Which datasets in that publication does this sheet apply to?:
This sheet applies to the complete dataset of the study reported in Chapter 4 of the dissertation.

3. Information about the files that have been stored
==========
3a. Raw data

The raw data consist of the peer feedback messages that were used for content analysis. Additionally, there is also data on the segmentation and coding process into the different content categories, which were employed during content analysis.

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:
* On which platform are the raw data stored?
 -[x] researcher PC
 -[x] research group file server

-[x] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

3b. Other files
---
* Which other files have been stored?
-[x] file(s) describing the transition from raw data to reported results. Specify: The content analysis coding scheme to examine the peer feedback content quality.
-[x] file(s) containing processed data. Specify: All the gathered data was cleaned, aggregated and restructured for analysis using SPSS.
-[x] file(s) containing analyses. Specify: SPSS-generated output was stored (i.e. output of preliminary analyses as well as output of the main analyses regarding the research questions).
-[ ] files(s) containing information about informed consent
-[ ] a file specifying legal and ethical provisions
-[ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
-[ ] other files. Specify: ...

* On which platform are these other files stored?
-[x] individual PC
-[x] research group file server
-[ ] other:
* Who has direct access to these other files (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
- name:
- address:
- affiliation:
- e-mail:

% Data Storage Fact Sheet
% Study 4_Chapter 5
% Author: Mario Gielen
% Date: March, 24, 2016

1. Contact details
==========
1a. Main researcher

- name: Mario Gielen
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: mario.gielen@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Bram De Wever (Supervisor PhD Project)
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: bram.dewever@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education, 88*, 370-386.

* Which datasets in that publication does this sheet apply to?:
This sheet applies to the complete dataset of the study reported in Chapter 5 of the dissertation.

3. Information about the files that have been stored
==========
3a. Raw data

The raw data consist of the peer feedback messages and of the draft and final version of the wiki product, which were measured with scoring rubrics.

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:
* On which platform are the raw data stored?
 -[x] researcher PC
 -[x] research group file server

-[x] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

3b. Other files
---
* Which other files have been stored?
-[ ] file(s) describing the transition from raw data to reported results. Specify: The rubrics to score the quality of the peer feedback messages and the wiki product.
-[x] file(s) containing processed data. Specify: The data was cleaned and aggregated in SPSS for analysis and restructured for multilevel analysis using MLwiN).
-[x] file(s) containing analyses. Specify: all MLwiN 2.23-generated model outputs (i.e. output of priliminary analyses as well as output of the main analyses regarding the research questions) were stored as .wsz files.
-[ ] files(s) containing information about informed consent
-[ ] a file specifying legal and ethical provisions
-[ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
-[ ] other files. Specify: ...

* On which platform are these other files stored?
-[x] individual PC
-[x] research group file server
-[ ] other:
* Who has direct access to these other files (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
- name:
- address:
- affiliation:
- e-mail:

% Data Storage Fact Sheet
% Study 5_Chapter 6
% Author: Mario Gielen
% Date: March, 24, 2016

1. Contact details
==========
1a. Main researcher

- name: Mario Gielen
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: mario.gielen@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Bram De Wever (Supervisor PhD Project)
- address: Henri Dunantlaan 2 - 9000 Ghent - Belgium
- e-mail: bram.dewever@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Gielen, M., De Wever, B. & Voet, M. (Accepted). The added value of structuring the role of both assessor and assessee during multiple feedback cycles in a wiki environment for students' peer feedback content. *Journal of Computer Assisted Learning.*

* Which datasets in that publication does this sheet apply to?:
This sheet applies to the complete dataset of the study reported in Chapter 6 of the dissertation.

3. Information about the files that have been stored
==========
3a. Raw data

The raw data consist of the peer feedback messages that were used for content analysis. Additionally, there is also data on the segmentation and coding process into the different content categories, which were employed during content analysis.

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:
* On which platform are the raw data stored?
-[x] researcher PC
-[x] research group file server

-[x] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

3b. Other files
---
* Which other files have been stored?
-[x] file(s) describing the transition from raw data to reported results. Specify: The content analysis coding scheme to examine the peer feedback content quality
-[x] file(s) containing processed data. Specify: The data was cleaned and aggregated in SPSS for analysis and restructured for multilevel analysis using MLwiN).
-[x] file(s) containing analyses. Specify: all MLwiN 2.23-generated model outputs (i.e. output of preliminary analyses as well as output of the main analyses regarding the research questions) were stored as .wsz files.
-[ ] files(s) containing information about informed consent
-[ ] a file specifying legal and ethical provisions
-[ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
-[ ] other files. Specify: ...

* On which platform are these other files stored?
-[x] individual PC
-[x] research group file server
-[] other:
* Who has direct access to these other files (i.e., without intervention of another person)?
-[x] main researcher
-[x] responsible ZAP
-[ ] all members of the research group
-[ ] all members of UGent
-[ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [x] NO
* If yes, by whom (add if multiple):
- name:
- address:
- affiliation: