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. This study illustrated how a practical instructional intervention in the feedback process can increase the potential impact of peer assessment and boost students' learning in higher education.

Keywords: computer-supported collaborative learning; peer assessment; scripting; peer feedback content; content analysis; wiki environment

1 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 divergent 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.

1.1 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. 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 perform their role as assessor and assessee proficiently (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).

1.2 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, 2014). 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.

1.3 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.*

1.4 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 the differential impact of a PFB template with a fluctuating structuring degree 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 that assessors give. In this way, 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.

2 Material and methods

2.1 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 compulsory introduction course in their program, accounting for 7 ECTS. This course was organised during the first semester of the academic year and more particular in the academic year 2012 – 2013 at the university of Ghent. Participating in the wiki-assignment was part of their curriculum requirements. Participants were randomly assigned to groups (n = 37) of maximum 4 to 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.

2.2 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.

2.3 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

2.4 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.

2.4.1 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).

2.4.2 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 peer feedback 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 peer feedback content quality, by concentrating on the peer feedback style, type, and focus. Table 2 represents and exemplifies the five coding categories: peer feedback style, verification type and focus, and elaboration type and focus.

2.4.3 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 peer feedback 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.

2.5 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*

3 Results

3.1 Descriptives

The descriptive results show significant differences between the three conditions regarding the number of segments per student [F (2,38)=67.149, p<.001, partial eta squared=.78]. In more detail, the descriptive results reveal 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.

3.2 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).

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

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).

3.3 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 $p < .05$

3.4 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).

3.5 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

3.6 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

4 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, 2015) 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.

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 for analysing 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 to the assessees if something is right

or wrong, but they also equally offer information on why this was right or wrong, in combination with ideas to improve their performance. Additionally, results indicated that students provide mostly positive comments, while the elaboration component consists almost equally out 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 for this reason 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 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 assessee in function of future performance (Carless, 2007).

5 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 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). 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*. doi: 10.1111/jcal.12096

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, *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. (2014). 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*. Advance online publication. Retrieved from http://dx.doi.org/10.1080/02602938.2014.898737.

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.