



HAL
open science

Peer feedback content and sender s competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?

Jan Willem Strijbos, Susanne Narciss, Katrin Dünnebier

► To cite this version:

Jan Willem Strijbos, Susanne Narciss, Katrin Dünnebier. Peer feedback content and sender s competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?. Learning and Instruction, 2010, 20, pp.291-303. hal-00703922

HAL Id: hal-00703922

<https://hal.science/hal-00703922v1>

Submitted on 4 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PEER FEEDBACK CONTENT AND SENDER'S COMPETENCE LEVEL IN
ACADEMIC WRITING REVISION TASKS:
ARE THEY CRITICAL FOR FEEDBACK PERCEPTIONS AND EFFICIENCY?

This is a post-print of an article submitted for consideration in the Learning and Instruction © 2010 Elsevier.

Personal use of this manuscript is permitted. Permission from Elsevier must be obtained for any other commercial purpose.

This article may not exactly replicate the published version, due to editorial changes and/or formatting and corrections during the final stage of publication. Interested readers are advised to consult the published version which can be found at:

<http://www.sciencedirect.com/science/article/pii/S0959475209000760>

doi:{10.1016/j.learninstruc.2009.08.008}

Please refer this manuscript as:

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.

Running head: Peer feedback content and sender's competence level

Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?

Jan-Willem Strijbos ^{a,*}, Susanne Narciss ^b, Katrin Dünnebier ^c

^a *Centre for the Study of Learning and Instruction, Institute of Education and Child Studies, Faculty of Social and Behavioral Sciences, Leiden University, P.O. Box 9555, 2300 RB, Leiden, the Netherlands*

^b *Centre for Psychology of Learning and Instruction, Department of Psychology, Technical University Dresden, Zellescher Weg 17, D-01069, Dresden, Germany*

^c *Centre for Educational Research and Teacher Education, Faculty of Educational and Social Sciences, University of Wuppertal, Gaußstrasse 20, D-42097, Wuppertal, Germany*

Abstract

Peer feedback content is a core component of peer assessment, but the impact of various contents of feedback is hardly studied. Participants in the study were 89 graduate students who were assigned to four experimental and a control group. Experimental groups received a scenario with Concise General (CGF) or Elaborated Specific (ESF) feedback by a high or low competent peer. ESF by a high competent peer was perceived as more adequate, but led to more negative affect. Students in CGF groups outperformed ESF groups during treatment. Groups with a low competent peer outperformed groups with a high competent peer during the posttest. Feedback perceptions and performance were uncorrelated.

Keywords: Peer feedback; Peer assessment; Feedback content; Feedback perceptions; Text revision

* Corresponding author. Tel.: +31 71 5274048; fax: + 31 71 5273619.
E-mail address: jwstrijbos@fsw.leidenuniv.nl (J. W. Strijbos)

1. Introduction

Many assessment researchers stress that feedback on student performance is a central part of so-called formative assessment approaches (Assessment Reform Group, 2002; Black & Wiliam, 1998), and peer assessment in particular (Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002). Topping (1998) defines peer assessment as «an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status» (p. 250). Typically, students judge a peer's performance with one or several rating scales, comments, or a combination. Although the impact of peer assessment on learning is widely voiced, the number of studies that investigated learning effects is scarce (Van Gennip, Segers, & Tillema, 2009; Van Zundert, Sluijsmans, & Van Merriënboer, this issue).

Characteristic of formative peer assessment is the strong reliance on the role and importance of feedback. However, it is well-known that feedback does not automatically lead to positive results (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Narciss, 2008; Shute, 2008). In addition, as students are not experts in a subject area, peer feedback is susceptible to variation. Students also doubt their own and peers' knowledge within a given subject area (Hanrahan & Isaacs, 2001), as well as their own and peers' skill to peer assess (Van Gennip, Segers, Tillema, this issue; Walker, 2001).

In the field of writing instruction, in particular English as a Second Language (ESL), there are several descriptive studies that have documented peer feedback variation. For example, Lockhart and Ng (1995) identified four different feedback stances: authoritative (focused on problems and errors in the text), interpretive (focused on a personal evaluation of the text), probing (focused on understanding the writers' intended meaning) and collaborative (focused on negotiating intended meaning of the text). Van den Berg (2003) compiled these four stances in two types of feedback – evaluative (authoritative and interpretive) and informative (probing and collaborative) – in the context of peer assessment of essays. Up to now, only few studies addressed the content of peer feedback in the context of peer assessment (Cho & MacArthur, this issue; Gielen, Peeters, Dochy, Onghena, & Struyven, this issue; Prins, Sluijsmans, & Kirschner, 2006).

There is also accumulating evidence that students' emotional state can mediate the impact of feedback on their performance (Shute, 2008). Many students express concerns about the fairness of peer assessment (Sluijsmans, Dochy, & Moerkerke, 1999) and students frequently voice that evaluation is a role of the teacher (Brown, Irving, Peterson, & Hirschfeld, 2009; Zhang, 1995). Furthermore, the perceived or actual ability of both assessor and assessee (Kali & Ronen, 2008; Lin, Liu, & Yuan, 2001) appears to affect peer ratings and might have an important effect on both acceptance of peer feedback and application during subsequent performance. The qualification "equal status students" in Topping's (1998) definition might be retained in the sense of age or class-level of students, but there are evidently individual differences that affect perceived status and may impact peer feedback perceptions and subsequent performance. The present study aimed to investigate whether and how content-related and social features of peer feedback affect feedback perceptions and performance.

1.1. Feedback content, feedback perceptions, and performance

In instructional contexts the term “feedback” refers to post-response information which informs the learners on their actual states of learning and/or performance, in order to help them detect if their state corresponds to the learning aims in a given context (Narciss, 2006, 2008). Depending on the actual state of learning, feedback can provide a variety of information. In case of no gap between the actual and intended state feedback can, for example, provide information that confirms goal achievement, correctness of a response, or the achieved level of performance. In case of small or large gaps it can provide more or less detailed information, which can be more or less specifically related to learning tasks or processes. Hence, a large variety of feedback types is used in instructional contexts.

To describe the feedback type variety systematically, several recent reviews and syntheses of research on feedback adopt a multidimensional view of feedback (Hattie & Timperley, 2007; Narciss, 2008; Shute, 2008). Narciss’s view stresses three main facets – feedback content, form, and function – that determine the quality of a feedback message (Narciss, 2006, 2008; Narciss & Huth, 2004; Shute, 2008). Combining these facets allows designing a large variety of feedback types, which might have differential effects on feedback perceptions and performance.

1.1.1. Feedback content

Widely investigated types of feedback are (a) simple feedback types providing outcome-related information, and (b) elaborated feedback types providing additional information besides outcome-related information. Narciss (2006, 2008) developed a content-related classification of feedback components which aims at providing a structured overview on simple and elaborated feedback components, and can be used for the systematic construction of feedback.

Simple feedback components are knowledge of performance, knowledge of result, and knowledge of the correct response. An elaborated feedback component is dependent on the elaborated information provided, which might address: (a) knowledge on task constraints (provides information on task rules, task constraints and task requirements), (b) knowledge about concepts (provides information on conceptual knowledge), (c) knowledge about mistakes (provides information on errors or mistakes), (d) knowledge on how to proceed (know-how) (provides information on procedural knowledge), and (e) knowledge on metacognition (provides information on metacognitive knowledge).

1.1.2. Feedback content and performance

The question of which feedback content is most efficient (i.e., which has the most beneficial effects on performance) has received much attention in prior feedback research. Within most of these feedback studies, the issues of how learners perceive feedback content, and how the perceptions relate to performance have not been addressed explicitly. Yet, several authors have emphasised the ‘mindful processing’ of feedback as a critical factor for feedback efficiency (Kluger & DeNisi, 1996; Narciss, 2008; Poulos & Mahony, 2008). Unfortunately, the results of a large body of feedback research are mixed. Only some studies support the commonsense assumption that elaborated and specific feedback affects performance more positively than concise general feedback (see

Hattie & Timberley, 2007; Mory, 2004; Narciss, 2008; Shute, 2008). This assumption is reflected in results of interview and questionnaire studies on writing instruction, that is, students perceive feedback providing elaborated and specific advice more positively than short evaluative feedback (Arndt, 1993; Straub, 1997).

1.1.3. Peer feedback content, feedback perceptions, and performance

The issue of whether and how various peer feedback contents affect feedback perceptions and performance has been mostly investigated in exploratory studies with small sample sizes. For example, Tsui and Ng (2000) found that students perceived teacher and peer feedback as helpful if it provided specific and elaborated revision suggestions. If so, students incorporated it in their revisions, whereas they ignored any feedback that was either not specific enough and/or did not explain the problems. In addition, the perceptions of peer feedback – if measured – is commonly measured in terms of the single dimension “usefulness”, after the feedback has been applied and/ or at the end of the task (Kwok, 2008).

1.2. Feedback source, feedback perceptions, and performance

Besides feedback content, the source or sender providing the feedback is a crucial factor for the efficiency of feedback. In organisational contexts five sources can be distinguished, namely the formal organisation, the supervisor, the co-workers, the task, and one’s own self (Andrews & Kacmar, 2001; Greller & Herold, 1975). In instructional contexts there are also at least five feedback sources, namely the teacher, peer, parents, book or computer-based environment (Hattie & Timperley, 2007), and the task (if the learning goals and criteria for goal attainment are defined in a way that learners can self-assess their learning). Depending on the source’s characteristics, feedback content might be perceived as less useful or less credible, and affect task completion or learning differentially.

1.2.1. Feedback source and feedback perceptions

Characteristics of the feedback source have been widely discussed as critical factor for feedback perceptions and acceptance in organisational psychology (Greller & Herold, 1975; Ilgen, Fisher, & Taylor, 1979; Kluger & DeNisi, 1996; Leung, Su, & Morris, 2001). Giffin (1967) states there are at least five dimensions of a feedback source that influence a source’s credibility – expertise, reliability, intentions toward the receiver, dynamism, and personal attraction. Ilgen et al. (1979) consider expertise as one of the most important factors for feedback acceptance. Expertise of the feedback source is expected to depend on such factors as training, experience, competence level, and familiarity with the task domain (Birnbaum & Stegner, 1979).

In general, feedback from a person with a high level of expertise is assumed to be perceived as more positive than from a person with low expertise. The study by Klein, Kraut, and Wolfson (1971) revealed that satisfaction with feedback is influenced by the perception of the source’s familiarity with the work unit: feedback from a person with low familiarity was perceived less positive than feedback from a person with high familiarity. Similarly, Halperin, Snyder, Shenkel, and Houston (1976) found that perceived feedback credibility depended on a source’s expertise. Finally, expertise of the

source has been found to influence intrinsic motivation (Cusella, 1982).

1.2.2. Peer feedback source, feedback perceptions, and performance

In the field of writing instruction and language learning the issue of feedback perceptions and the source is reflected in a controversial debate on peer versus teacher feedback. Several researchers have found that students perceive peer feedback less favourably than teacher feedback because a peer is perceived as less competent (Leki, 1991; Lockhart & Ng, 1993; Zhang, 1995). Others advocate that most students welcome and use peer feedback as one kind of feedback to improve their writing (Jacobs, Curtis, Braine, & Huang, 1998; Tsui & Ng, 2000). Recent studies indicate that peer feedback can be associated with a larger degree of student autonomy (Yang, Badger, & Yu, 2006) and teacher feedback can result in passive and dependent learners (Lee, 2008).

The issue of sender's competence level has so far been addressed indirectly in most peer assessment research by comparing student and teacher marks (Falchikov & Goldfinch, 2000). Irrespective of a high correlation between student and teacher marks, students voice concerns about the fairness and usefulness of a peer assessment (Brown et al., 2009; Sluijsmans et al., 1999). Similar to studies on writing instruction most students value peer feedback, but prefer teacher feedback because they consider them more competent.

Finally, there is also contradictory evidence as to the reliability and validity of self-assessment compared to peer assessment (Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006). If students have to self-assess their learning they may experience less stress and discomfort, as compared to when their work is reviewed by a peer or a teacher. In a perceived unfair situation (real or imagined), a negative reaction will be stronger compared to a non-threatening context (Miedema, 2004), thus, the acceptance of peer feedback (irrespective of accuracy) becomes less likely.

In summary, there is need to investigate in more detail (a) how the sender's competence level affects peer-feedback perceptions and (b) how these peer-feedback perceptions relate to feedback efficiency.

1.3. The present study

As the literature review spans several research domains the terminology used fluctuates and, therefore, in the present study ability, competence, and expertise are treated as interchangeable; for clarity we use "competence" (professionals are deemed experts, but students are not experts in educational settings). Also, the peer feedback is always sent by a person, which is why we use the term "sender".

The present study aimed to investigate in an experimental design the impact of the contents of peer feedback (concise general vs. elaborated specific feedback) and sender's competence level (low vs. high) on peer-feedback perceptions and performance in the context of academic writing. More specifically, the study had three aims: (a) to develop and test a multidimensional questionnaire measuring feedback perceptions; (b) to investigate the possible interaction of feedback content (concise general vs. elaborated specific) and sender's competence level (high vs. low) and its effect on feedback perceptions and performance; and (c) to explore whether and how the effects of feedback content, feedback sender's competence level, and feedback perceptions relate to

performance.

To meet these aims a study with four experimental groups and one control group was conducted. The experimental groups varied on the factors Feedback Content, namely Concise General Feedback (CGF) versus Elaborated Specific Feedback (ESF), and Sender's Writing Competence Level (high vs. low). The control group received no feedback.

To investigate the sender's competence level in a controlled manner four scenarios, one for each experimental group, were used. Prior research testing the convergence of reactions to real situations versus scenarios has shown that persons react almost identically to scenarios as they do to real situations (Robinson & Clore, 2001).

1.3.1. Research questions and Hypotheses

Prior research on effects of feedback content and sender characteristics, more specifically sender's competence level, reveals that students' perception of feedback is very important. However, prior findings on the effects of various feedback contents on performance are mixed.

The first research question of the present study regarded the measurement of feedback perceptions. It was hypothesized that feedback perceptions would be adequately captured with the theoretically developed multidimensional Feedback Perception questionnaire (Hypothesis 1).

The second research question regarded the effects of the content of peer feedback and sender's competence level on the perceptions of peer feedback and subsequent performance. Based on the findings that students prefer elaborated feedback, it was expected that elaborated specific feedback as compared to concise general feedback will (a) be perceived as more adequate, (b) lead to more willingness to improve, and (c) lead to more positive affect (Hypothesis 2a). Furthermore, prior studies have revealed that students prefer teacher feedback over peer feedback because they consider the teacher to be more competent. Thus, it was assumed that feedback from a high competent sender as compared to a low competent sender will (a) be perceived as more adequate, (b) lead to more willingness to improve, and (c) lead to more positive affect (Hypothesis 2b). Moreover, combining Hypotheses 2a and 2b, it was hypothesized that elaborated feedback by a high competent peer would be perceived as most adequate, whereas concise feedback by a low competent peer would be perceived as the least adequate (Hypothesis 2c). As regards the effects of peer feedback content and sender's competence level on performance, it was expected that peer feedback groups will outperform the control group, as in many experimental feedback studies any feedback has proved to be better than no feedback (Hypothesis 3a). Based on research comparing the effects of teacher versus peer feedback, it was assumed that feedback by a high competent peer affects performance more positively than feedback by a low competent peer (Hypothesis 3b).

The third research question regarded the relation between feedback perceptions and feedback effects on performance. Prior research relating feedback perceptions to feedback effects on performance is missing; hence no specific hypothesis was formulated.

2. Method

2.1. Design

The present study investigated how feedback content and sender's competence level affect feedback perceptions and performance. Students were assigned to four experimental and a control group. Students in the experimental groups received a scenario in which a fictional student received concise general (CGF) or elaborated specific (ESF) feedback by a fictional peer whose competence was high (H) or low (L). The four experimental groups were the CGF high competent group (CGF-H), the CGF low competent group (CGF-L), the ESF high competent group (ESF-H), and the ESF low competent group (ESF-L).

A two-way factorial design with the five groups (four experimental and a control group) as between subjects factor and three phases (pretest, treatment and posttest) as within subjects factor was applied. Performance on text revision tasks was the dependent variable.

2.2. Participants

The participants were 89 graduate teacher training and psychology students from the Technical University Dresden. There were 68 female and 21 male students, and their age ranged from 20 to 40 years ($M = 24.71$, $SD = 4.14$). Participation was part of their curriculum requirements and they received no financial compensation.

Participants were randomly assigned to one of five research groups. There were 18 participants in each experimental group and the control group, except for the ESF low competent group which comprised 17 participants.

2.3. Materials

2.3.1. Task

The study was situated in the context of academic writing instruction. Participants were asked to revise one text containing various errors during each phase (pretest, treatment, posttest) in view of text comprehension criteria (simplicity, structure, conciseness, stimulating; see Appendix A). The focus on text comprehension criteria was selected because revisions on content-related level would have required specific knowledge on the topic of each text. The pretest was on "Rediscovered or false memories" (adapted from Pezdek, 2001). The posttest text was on "How speech can guide thought: linguistic relativism and linguistic determinism" (adapted from Slobin, 2001). Both texts have been used in research on academic writing (Proske, 2006). Each of these texts was modified in such a way that it contained typical errors with regard to the text comprehension criteria (i.e., very long sentences, technical terms without explanations, missing subheadings). The treatment text on "Foundations of sexuality" and the feedback were adopted from Langer, Schulz von Thun, and Tausch (1999, pp. 44-56). The first error always served as an example and indicated how students were supposed to identify errors. The pretest contained 13, the treatment 24 and the posttest 29 errors.

2.3.2. *Feedback scenarios*

There were four scenarios. Feedback content was either CGF or ESF and provided by a high or low competent peer. The CGF content provided solely remarks on knowledge of results, whereas the ESF content provided the position and error type, as well as information on how to proceed (Appendix B shows error types in terms of the feedback components classification by Narciss, 2006, 2008). The sender's competence level (high vs. low) was mentioned three times in the treatment phase with a phrase such as "The following feedback is provided by a peer who, thus far, has achieved high/ low performances in writing".

2.3.3. *Manipulation check*

Prior to the third text a manipulation check was performed with yes or no questions, such as "The feedback was provided by a peer who, thus far, had high quality written performances". There were three manipulation check questions.

2.4. *Measures*

2.4.1. *Feedback Perceptions questionnaire*

In the treatment phase the participants were asked to consider the feedback as if they had received the feedback themselves, and indicate how they perceived the feedback in terms of fairness, usefulness, acceptance, willingness to improve, and affect. Questionnaire items were measured on a 10 cm bi-polar scale from 0 (fully disagree) to 10 (fully agree). All five scales of the questionnaire were developed for this study. Four of the scales consisted of three items: (a) Fairness (e.g., "I would consider this feedback fair"; Cronbach's $\alpha = .80$); (b) Usefulness (e.g., "This feedback would provide me a lot of support"; Cronbach's $\alpha = .82$); (c) Acceptance (e.g., "I would dispute this feedback"; Cronbach's $\alpha = .69$); and (d) Willingness to Improve (e.g., "I would be willing to improve my performance"; Cronbach's $\alpha = .82$). Affect was measured with six items (e.g., "I would feel offended if I would have received this feedback"; Cronbach's $\alpha = .81$). Negative phrased items were recoded so that the scale measured positive affect.

2.4.2. *Knowledge of text comprehension criteria*

At the start of the pretest students were asked to write down what they considered to be criteria for a readable and understandable text based on their personal writing experience. The instruction of the treatment phase included the formal criteria (see Appendix A) for text comprehension (Langer et al., 1999). Knowledge of text comprehension criteria was measured at the beginning of the posttest as follows: "Please list the text comprehension criteria and distinguish main and sub-criteria". In the pretest and posttest knowledge of text comprehension criteria was operationalised as (a) the number of correctly reproduced sub-criteria (16 overall), (b) the number of sub-criteria that were correctly ordered to the main criteria, and (c) the number of correctly reproduced main criteria (4 overall).

Sub-criteria and their ordering to main criteria were scored according to clear scoring rubrics. For example, the sub-criterion "short and simple sentences" has to be assigned to the main criterion "Simplicity". It can be addressed by various statements, such as "avoid complex sentences", "use simple sentence structures" or "use short rather

than long sentences”. All these statements and all their variations were scored as correct statements for the sub-criterion “short and simple sentences”. Students’ knowledge of sub-criteria and ordering to main criteria was computed as the sum of correctly (1 point) and incorrectly (0 point) stated sub-criteria and orderings. These scoring rubrics were piloted with a small subset of data (pretest data of 10 participants) and proved to be unambiguous.

2.4.3. Revision performance

During the pretest, treatment, and posttest students had to revise a text. Each of these texts contained typical errors with regard to the text comprehension criteria (i.e., very long sentences, technical terms without explanations, missing subheadings). These errors were intentionally included into the text by one of the authors.

The instructions were as follows: “Please revise the following text to enhance its comprehensibility. The corrections can involve a word, sentence or multiple sentences. Please identify the text fragments that are difficult to comprehend with brackets and state which text comprehension criterion is violated. Please write down your revision proposal for each identified fragment. An example is provided”. Text revision performance was operationalised in terms of (a) the number of correctly identified errors, (b) the number of incorrectly identified errors (false alarms), (c) the number of unnoticed errors, (d) the number of correctly explained errors, (e) the number of correct revision proposals to correctly identified errors, and (f) the proportion of correctly identified errors relative to the total number of errors in a given text.

For each of the indicators (a) thru (e) frequencies were computed. Indicator (f) was computed as the proportion of indicator (a) relative to the total number of errors. As the text comprehension errors had been included intentionally, the scoring of the revision performance indicators was unambiguous.

2.4.4. Time on task

We measured students’ time on task during revision, as this is an important indicator for performance (Van Gog et al., 2008). Students manually recorded the time when starting the text revision and upon finishing it.

2.5. Procedure

During the pretest all participants were asked to write down what they considered to be criteria for a readable and understandable text, and then revised a text that contained several errors that made the text hard to read and understand. During the treatment the participants studied a scenario, which consisted of a text revised by a fictional student and feedback that this fictional student received by a fictional peer, as the revised text still included text comprehension errors. Feedback content (CGF vs. ESF) and sender’s competence level (high vs. low) were manipulated. The participants were asked to rate the fictional feedback – as if had they themselves received this feedback – in terms of fairness, usefulness, acceptance, willingness to improve and affect. Next, they were asked to study the text comprehension criteria, which were given in the scenario, and applied the feedback to a second revision of the ‘revised text’. The control group received the same criteria, revised the same text, but received no feedback. During the posttest the

participants were asked to describe the text comprehension criteria that were presented during the treatment and revised a third text.

The data was collected during three educational psychology seminars. In each seminar the experiment lasted ninety minutes (pretest, treatment, posttest; each phase lasted 30 minutes). One student did not hand in the posttest materials.

3. Results

Prior to the analyses, first, the distribution assumptions, decisions made in relation to missing values, and checks for seminars are addressed. Then, the manipulation check is reported.

3.1. Data inspection

We first checked distribution assumptions. The standardised skewness and kurtosis were outside the +3 to -3 range for pretest revision performance (4.00 and 3.33, respectively), and kurtosis was outside this range for posttest sub-criteria (3.08). Examination of all variables revealed 15 univariate outliers. These were checked for influence on mean, skewness and kurtosis. Outliers were adapted to the nearest highest score or in case of strong deviations one unit above the highest score in order to retain to a degree the difference with the nearest highest score (Tabachnick & Fidell, 2001). Two outliers on pretest revision performance were adapted to the nearest score, because it strongly reduced skewness and kurtosis. One outlier on the posttest sub-criteria was adapted one unit above the nearest highest score, as it reduced kurtosis strongly. No bivariate or multivariate outliers were found.

Four variables had more than 5% missing values: posttest sub-criteria (10.1%), time on task during the pretest (31.5%), time on task during the treatment (6.7%) and time on task during the posttest (7.9%). We used the EM (Expectation Maximisation) procedure for imputation (Musil, Warner, Yobas, & Jones, 2002). Subsequently, we checked whether this imputation influenced the outcomes of all planned analyses. No discrepancies with respect to significant findings were observed. Hence, all analyses further on are reported using the imputed dataset with no missing values. We also checked for systematic differences between the three seminars. As this was not the case we collapsed the data.

3.2. Manipulation check

Both manipulation check questions on feedback content appeared to be ambiguous in retrospect, but the question on the sender's competence level revealed that at least 69% of the participants (up to 89%) in the experimental groups referred to the competence level correctly (overall 80% answer correct).

3.3. Feedback Perceptions questionnaire

The means for the five scales of the questionnaire were computed, as well as the correlations between them. There was common variance to a considerable degree

reflected by moderate to high correlations (see Table 1), particularly between fairness, usefulness, and acceptance.

 Insert Table 1 about here

For this reason all items of the questionnaire were, then, entered to a principal component analysis with an oblimin rotation (see Table 2). Four factors were revealed (loadings > .400), namely Willingness to Improve ($R^2 = .09$, Cronbach's $\alpha = .82$), Positive Affect ($R^2 = .11$, Cronbach's $\alpha = .90$), and Negative Affect ($R^2 = .15$, Cronbach's $\alpha = .83$). Fairness, Usefulness, and Acceptance formed conceptually a fourth factor labelled Perceived Adequacy of Feedback (PAF, Cronbach's $\alpha = .89$) explaining 35.7% of all variance. The factor correlations were modest.

 Insert Table 2 about here

Following the outcome of the principal component analysis, it was decided that the PAF scale should be used in further analyses. Furthermore, although positive and negative affect constituted two distinct factors, Cronbach's alpha for all six items was sufficient (.81), thus it was decided that the entire Affect scale be used as a single measure reflecting positive affect after conversion of the scoring of the negative affect items. In summary, in further analyses the following three scales of the Feedback Perceptions questionnaire were used: (a) PAF, (b) Willingness to Improve (WI), and (c) Affect (AF).

3.4. Impact of feedback content and sender's competence level on feedback perceptions

Pearson correlations revealed that PAF was moderately associated with WI ($r = .46$, $p < .01$) and AF ($r = .38$, $p < .01$). Spearman correlations were, then, computed for PAF and WI within each group. Specifically, $\rho(18) = .50$, $p < .05$, for CGF-H; $\rho(18) = .63$, $p < .01$, for CGF-L; and $\rho(17) = .69$, $p < .01$, for ESF-L; no correlation was found for ESF-H. When split by group there were no correlations between PAF and AF, and between WI and AF.

A 2 x 2 MANOVA was performed with feedback content and sender's competence level as the independent variables, and PAF, WI and AF as dependent variables. Levene's test indicated homogeneity of variances for all dependent variables (PAF = .58, WI = .50, and AF = .48). The results showed a multivariate interaction between feedback content and sender's competence level, Pillai's trace = .13, $F(3, 65) = 3.24$, $p = .028$, partial $\eta^2 = .13$. There was no multivariate main effect for feedback content, Pillai's trace = .11, $F(3, 65) = 2.57$, $p = .062$, but the multivariate main effect for sender's competence level was significant, Pillai's trace = .11, $F(3, 65) = 2.80$, $p = .047$, partial $\eta^2 = .11$. Table 3 provides means and standard errors for PAF, WI and AF by group.

 Insert Table 3 about here

Follow-up univariate analyses were computed with the Bonferroni correction set

at alpha level $< .05/3 = .016$. There was a significant interaction for AF between feedback content and sender's competence level, $F(1, 67) = 6.87$, $p = .011$, partial $\eta^2 = .09$ (see Figure 1). There was also a significant main effect for feedback content on PAF, $F(1, 67) = 6.49$, $p = .013$, partial $\eta^2 = .09$, and for sender's competence level on PAF, $F(1, 67) = 6.24$, $p = .015$, partial $\eta^2 = .08$.

As recommended by Field (2005), a discriminant analysis revealed that there was one significant common variate, Wilks's lambda = .695, $df = 9$, $p = .004$. It was found that PAF was a strong and positive contributor to the variate (.73), AF was a moderate negative contributor (-.24), and WI's contribution was small (.09). The group centroids revealed that the ESF-H group perceived the feedback differently (opposite sign compared to the other groups) in terms of a combined effect of PAF, AF and WI (CGF-H = -.16, CGF-L = -.44, ESF-H = .83 and ESF-L = -.24). Specifically, feedback in the ESF-H group was perceived as the most adequate, but also led to negative affect. The ESF-H group was the only group with a large discrepancy between PAF and AF. Figure 1 illustrates the multivariate interaction.

 Insert Figure 1 about here

3.5. *Impact of feedback content and sender's competence level on performance measures*

The impact of feedback content and sender's competence level was investigated with regard to knowledge of text comprehension criteria and revision performance by time on task. There were modest correlations between indicators for knowledge of sub-criteria and their classification to main criteria. Thus, it was decided to include only the number of correctly stated sub-criteria into further analyses.

With respect to revision performance indicators we noticed that nearly all, except for the number of false alarms, correlated strongly ($r = .92$ to 1.00). Therefore, it was decided that all analyses be performed only on the performance indicator "number of correct revision proposals to correctly identified errors".

 Insert Table 4 about here

Table 4 shows the means and standard errors for revision performance (i.e., the number of correct revision proposals to correctly identified errors), knowledge of sub-criteria and time on task. The pretest and posttest knowledge of sub-criteria were positively – but moderately – associated, $r(89) = .27$, $p < .01$, which was also the case for revision performance on the pretest and posttest, $r(89) = .39$, $p < .01$.

There was a moderate positive correlation for time on task and revision performance in each phase. Specifically, for pretest, $r(89) = .34$, $p < .01$; for treatment, $r(89) = .34$, $p < .01$; and for posttest, $r(89) = .41$, $p < .01$. Given these correlations, "revision performance by time on task" was a more adequate measure and therefore a ratio was computed of revision performance by time (in minutes) on task as a function of phase and group (see Table 4).

3.5.1. Impact of feedback content and sender's competence level on knowledge of sub-criteria

A repeated measures ANOVA was then performed on knowledge of sub-criteria scores for all five groups. The results revealed a main effect for the repeated measure knowledge of sub-criteria, Pillai's trace = .09, $F(1, 84) = 7.93$, $p = .006$, partial $\eta^2 = .09$. Table 4 shows that knowledge of sub-criteria increased from pretest to posttest in the ESF-H, ESF-L and control group. For the CGF-H and CGF-L groups there was no increase compared to their pretest revision performance (which was higher than the other groups, but not statistically significant).

3.5.2. Impact of feedback content and sender's competence level on revision performance by time on task

Next, a repeated measures ANOVA was performed on revision performance by time on task scores. Mauchly's test indicated a violation of the assumption of sphericity, $\chi^2(2, N = 89) = 10.30$, $df = 2$, $p = .006$, and the degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity ($\epsilon = .96$). The results revealed an interaction between group and revision performance by time on task, $F(7.6, 160.82) = 2.39$, $p = .02$, partial $\eta^2 = .10$. Paired contrasts were, then, performed to break down the interaction. These revealed a quadratic interaction, $F(4, 84) = 3.19$, $p = .017$, partial $\eta^2 = .13$, for group and revision performance by time on task.

The results showed that revision performance by time on task increased from pretest to treatment to posttest, but the pattern over phase was different for each group. Revision performance by time on task in the CGF-L, ESF-H, and ESF-L groups increased from one phase to the next, whereas for the CGF-H group revision performance by time on task stayed constant from the treatment to the posttest phase and for the control group it decreased slightly in the posttest compared to the treatment. The increase from treatment to the posttest was highest for the ESF-L group. Figure 2 illustrates the interaction between revision performance by time on task and group.

 Insert Figure 2 about here

Planned contrasts were conducted to investigate whether mean revision performance by time on task in the experimental groups differed from that in the control group. The results showed that the mean revision performance by time on task of the ESF-H group was significantly lower than that of the control group (the difference in means was -0.07 with a standard error of 0.03, $p = .021$). When applying the Bonferroni correction, set at alpha level $< .05/4 = .012$, this difference was no longer significant.

There was also a main effect of phase on revision performance by time on task, $F(1.9, 160.8) = 56.48$, $p < .000$, partial $\eta^2 = .40$. The contrast revealed a quadratic interaction for revision performance by time on task, $F(1, 84) = 6.39$, $p = .013$, partial $\eta^2 = .07$. Pairwise comparisons (Bonferroni adjusted) revealed that revision performance by time on task in the treatment ($M = 0.27$, $SD = 0.16$) and posttest ($M = 0.32$, $SD = 0.18$) were significantly higher than in the pretest ($M = 0.12$, $SD = 0.08$). The increase from treatment to posttest was marginally significant ($p = .055$) and accounted for the quadratic interaction.

Finally, differences between the experimental groups within the treatment and

posttest phase were investigated. A 2(feedback content) x 2(sender's competence level) ANOVA revealed a significant difference for revision performance by time on task in the treatment phase. Specifically, students in the CGF groups outperformed students in the ESF groups, $F(1, 67) = 8.23$, $p = .005$, partial $\eta^2 = .11$. We then specifically checked whether this difference could be attributed to the errors and elaborations mentioned in the ESF – which the CGF groups did not receive – but this was not the case, $F(1, 67) = 0.54$, $p = .463$. A 2(feedback content) x 2(sender's competence level) ANOVA for revision performance by time on task in the posttest revealed that students who experienced fictional feedback by a low-competence peer outperformed students who experienced feedback by a high-competence peer, $F(1, 67) = 4.72$, $p = .033$, partial $\eta^2 = .07$.

3.6. The relation between feedback perceptions and performance measures

The relation between feedback perceptions and performance measures was investigated through correlational analyses. Knowledge of sub-criteria during the pretest correlated negatively with PAF, $r(71) = -.37$, $p < .01$, but when split by group the Spearman correlation was only present for the ESF-H group, $\rho(18) = -.66$, $p < .01$. In addition, WI correlated positively with time on task during treatment, $r(71) = .35$, $p < .01$, but when split by group the Spearman correlation was only present for the ESF-H group, $\rho(18) = .48$, $p < .05$. There were no correlations between PAF, WI, AF and revision performance by time on task in any phase.

4. Discussion

The present study investigated how feedback content and sender's competence level affect feedback perceptions and performance (knowledge of text comprehension criteria and revision performance by time on task). Students were assigned to four experimental and a control group. Students in the experimental groups received a scenario in which a fictional student received concise general (CGF) or elaborated specific (ESF) feedback by a fictional peer whose competence was high (H) or low (L).

4.1. Measuring feedback perceptions

First, a multidimensional questionnaire was developed to measure feedback perceptions in terms of fairness, usefulness, acceptance, willingness to improve, and affect. The results revealed distinct scales for perceived adequacy of feedback (PAF; consisting of fairness, usefulness and acceptance), willingness to improve (WI), and affect (AF). This finding indicates that (at least) three dimensions of feedback perceptions can be discerned (Hypothesis 1 was confirmed) and these might be differentially related to revision performance. Moreover, our data showed that other dimensions, namely willingness to improve and affect, are important besides usefulness. Finally, it appears that students do not distinguish between fairness, usefulness and acceptance.

4.2. Effect of feedback content and sender's competence level

The manipulation check showed that a strong majority of the participants could identify the sender's competence level. Perhaps ESF is automatically perceived as being from a more competent person. Consequently, ESF by a high competent peer might be perceived similarly as feedback by a teacher (Tsui & Ng, 2000). Thus, if the students are used to feedback by a more competent sender, this may shape how they perceive the peer feedback and as a result they may not have processed the manipulation consciously. If so, feedback content might be more important for feedback perceptions compared to the competence level of a peer, and this notion supports the role of instruction (Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, THIS ISSUE) on peer feedback quality. Nevertheless, there might be compensatory mechanisms that mediate between feedback content and sender's competence level.

Complicated interaction effects of feedback content and sender's competence level for both feedback perceptions and revision performance by time on task were found. On the one hand, ESF was perceived as more adequate than CGF, and feedback by a high competent peer was perceived as more adequate than feedback from a low competent peer (Hypothesis 2a and 2b were both partially confirmed). Yet, elaborated specific feedback by a high competent peer was perceived as more adequate, but also resulted in more negative affect (Hypothesis 2c was partially confirmed). Apparently the peer's competence matters when peer feedback is elaborated and specific, and as such it is not perceived similar to teacher feedback. The response to feedback by a particular peer could be more extreme due to students' relative status, or ESF by a high competent peer might be perceived as a threat to the self (Miedema, 2004).

Regarding revision performance by time on task, the ESF-H group was the only experimental group that differed from the control condition – their revision performance by time on task was overall lower and it improved more slowly over phase (Hypothesis 3a was rejected). This finding stresses a differential experience of the ESF-H group compared to the other groups. However, the revision performance by time on task difference of the ESF-H group disappeared when the Bonferroni correction was applied – but it should be noted that this correction is conservative and contains a risk of an inflated Type II error in the pursuit of reducing the risk of a Type I error. The disappearing difference might also be explained by a lack of statistical power; hence we leave it to reader to decide how to treat this finding best. Next, groups with feedback by a low competent peer outperformed groups with feedback by a high competent peer during the posttest (Hypothesis 3b was rejected). Finally, the finding that CGF groups outperformed ESF groups during the treatment, contributes to the overall mixed effects of feedback content on revision performance by time on task. It might be that the students receiving elaborated feedback by a high competent peer may have relied more on this feedback and thus were less inclined to actively search and correct the text comprehension errors. Elaborated feedback by a high competent person may render the student passive and dependent on the feedback (Lee, 2008). Clearly, this issue deserves further investigation.

4.3. Relation between feedback perceptions and performance measures

Direct relations between feedback perceptions and performance measures were not found. This is not in line with the commonsense assumption that the elaborated feedback would be perceived as more positive and, thus, would lead to better performance. If the feedback were experienced as negative, it would intuitively impede the performance, and the opposite for positive perception. It is possible that our feed forward (by providing criteria to all groups at the start of the treatment phase) may have led our participants to self-assess the text and that the external feedback was redundant. Since they did not necessarily need the feedback, it may be no wonder that feedback perceptions were not directly related to revision performance, and that the experimental groups did not outperform the control group. In further studies we aim to examine the same scenario without providing the feed forward information. Even though the sender's competence level plays a role for feedback perceptions, it does not affect the effectiveness of feedback – as long as all students are provided with criteria on which the feedback is based. Nevertheless, apparently other factors may influence this relation. For example, further research could investigate what constitutes student status and how sender characteristics interact in relation to feedback perceptions.

4.4. Methodological limitations

On the one hand a scenario technique might be considered artificial, but on the other hand prior research (Robinson & Clore, 2001), comparing reactions to scenarios with reactions to real situations, showed that there is a high degree of correspondence in these reactions. Furthermore, in a real peer-feedback setting other interpersonal variables (for example familiarity and gender) might have masked the effect of the sender's competence level manipulation. Investigating the interplay of interpersonal variables with various feedback contents deserves further research. In addition, classroom studies can complement the outcomes of experimental studies and provide us further insights on how students process feedback (perceptions) and the effect on performance (e.g., the role of "mindful processing").

The time on task measure was recorded manually, which may have induced some bias, and, if available, a computer-based environment is preferred for future studies. The time on task measure during the treatment phase also covered the questions on the perceptions of feedback, but the differences in time on task are unlikely due to answering these questions. However, we acknowledge that the measure could have been more precise and we will adapt the experimental material in our future studies. Furthermore, the number of errors in each text could be kept constant. Finally, the feedback during the treatment was taken from Langer et al. (1999) and was assumed to be relevant and helpful. A preliminary study could have been conducted to check the quality of feedback before the manipulation.

Our feedback focused on surface features and not deep features (Hattie & Timperley, 2007) and it may have contributed to the performance findings, that is, that the experimental groups did not outperform the control group, as surface features are easier to correct than deep features. We selected text comprehension criteria because content-related revisions would have required more specific knowledge on the topic of a

text. Prior studies have shown that task difficulty might be a critical factor for feedback efficiency (Mory, 2004; Narciss, 2008), thus any generalisation of our results should be limited to the beginning level of writing, rather than the more general term academic writing. In all, it remains an open question for further research whether the effect on perceptions of feedback and performance would be different for more complex tasks.

The lack of differences between the experimental groups and the control group may also be explained by motivational problems. Some students reported feeling frustrated because they had to revise three texts. The posttest was also more difficult than the treatment text. Finally, the lack of differences between the treatment and posttest revision performance might be explained by the fact that the pretest activated prior knowledge, and might have raised student performance during the treatment, thus subsequently obscuring posttest differences.

4.5. Practical implications

The findings of the present study suggest that during peer assessment and peer feedback exercise the difference in competence level should not be too strong, because the students in the ESF-H group experience negative affect which apparently translates into less and a slower performance improvement. In addition, teachers should not assume that elaborated feedback works automatically better – although these are preferred they are not necessarily more efficient and may even render learners passive and dependent (Lee, 2008). Finally, there is a concern in the peer feedback literature that peer feedback by a low competent peer would be less effective, because it is perceived as less adequate. Yet, the findings show that feedback by a low competent peer is equally effective. Hence, teachers should not be worried about including low competent students in a peer assessment exercise with well-designed instructions.

Appendix A. Text comprehension criteria (main criteria and sub-criteria)

| Main criteria | Sub-criteria |
|---------------|---|
| Simplicity | Short and simple sentences, simple sentence structure Use of common words (few uncommon words, adjust words to audience) Explain jargon Concrete and clear |
| Structure | Subsections and structured Logical and correct order Well-organised (sections and indentation) Distinguish major and minor issues Red line |
| Conciseness | Only the essence/ importance Short, precise Focus on learning goal Every word is necessary |
| Stimulating | Exciting, interesting (examples, stimulating) A lot of variation Personal (catch reader's interest, everyday life) |

Appendix B. Concise General Feedback (CGF) and Elaborate Specific Feedback (ESF) components

| | Type of feedback | Feedback component |
|------------------------------------|--|--|
| Concise General Feedback (CGF) | | |
| Simplicity | The text is pretty complex and hard to understand. | KR |
| Structure | The structure is reasonable, but can be improved in areas. | KR |
| Conciseness | The text is rather long and long-winded. | KR |
| Stimulating | The text could be more stimulating. | KR |
| Elaborated Specific Feedback (ESF) | | |
| Simplicity | Partly a simple form of communication (e.g., “because otherwise these tribes and people would not be around anymore.”). | KR KH |
| | But mostly long sentences. Also uncommon words: driving comfort, ritual, expression. | KR KR + KM + KH (implicit) |
| Structure | The words “previously – meanwhile – nowadays” as markers provide a rough structure. | KR + KH |
| | Nevertheless, an introduction is missing that prepares the reader for this organisation. Also: the examples (car and homosexuality) are not really part of the main line of thought. It is hard to recognise what is actually an example for something. | KR + KM + KH (implicit) KR + KM + KH (implicit) |
| Conciseness | Both with respect to language usage and the content long-winded. | KR |
| | Language usage: “to a new opinion and view”; “it shows and expresses”; “based on the rich new knowledge, that came there to light”; “scientists and theoreticians”; “research and comparisons of the research”. Content: the Meanwhile section is nearly superfluous, the examples too extensive. | KM + KH (implicit) KM + KH |
| Stimulating | Examples/comparisons; further little stimulating. | KR + KH KM |
| Comprehensibility | Poor | KR |

KR = knowledge of result or response, KM = knowledge of error or mistake, and KH = knowledge on how to proceed (for more details, see Narciss, 2006, 2008).

References

- Andrews, M. C., & Kacmar, K. M. (2001). Confirmation and extension of the sources of feedback scale in service-based organizations. The Journal of Business Communications, *38*, 206-226.
- Arndt, V. (1993). Response to writing: Using feedback to inform the writing process. In M. N. Brock & L. Walters (Eds.), Teaching composition around the pacific rim: Politics and pedagogy (pp. 90-116). Clevedon, UK: Multilingual Matters.
- Assessment Reform Group. (2002). Assessment for learning: 10 principles. Retrieved 21 May, 2009, from <http://www.assessment-reform-group.org/CIE3.PDF>
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgments: Bias, expertise, and the judges's point of view. Journal of Personality and Social Psychology, *37*, 48-74.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. London: King's College, School of Education.
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. Learning and Instruction, *19*, 97-111.
- Cho, K., & MacArthur, C. (this issue). Student revision with peer and expert reviewing. Learning and Instruction.
- Cusella, L. P. (1982). The effects of source expertise and feedback valence on intrinsic motivation. Human Communication Research, *9*, 17-32.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of Educational Research, *70*, 287-322.
- Field, A. (2005). Discovering statistics using SPSS (2nd ed.). London: Sage.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (this issue). Improving the effectiveness of peer feedback for learning. Learning and Instruction.
- Greller, M. M., & Herold, D. M. (1975). Sources of feedback: A preliminary investigation. Organizational Behavior and Human Decision Processes, *13*, 244-256.
- Giffin, K. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. Psychological Bulletin, *68*, 104-120.
- Halperin, K., Snyder, C. R., Shenkel, R. J., & Houston, B. K. (1976). Effect of source status and message favorability on acceptance of personality feedback. Journal of Applied Psychology, *61*, 85-88.
- Hanrahan, J., & Isaacs, G. (2001). Assessing self- and peer assessment: The students' views. Higher Education Research and Development, *20*, 53-70.
- Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, *77*, 81-112.
- Ilgen, D. R., Fischer, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, *64*, 349-371.
- Jacobs, G. M., Curtis, A., Braine, G., & Huang, S.-Y. (1998). Feedback on student writing: Taking the middle path. Journal of Second Language Writing, *7*, 307-317.
- Kali, Y., & Ronen, M. (2008). Assessing the assessors: Added value in web-based multi-cycle peer assessment in higher education. Research and Practice in Technology Enhanced Learning, *3*, 3-32.

Klein, S. M., Kraut, A. I., & Wolfson, A. (1971). Employee reactions to attitude survey feedback: A study of the impact of structure and process. Administrative Science Quarterly, *16*, 497-514.

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.

Kwok, L. (2008). Students' perception of peer evaluation and teachers' role in seminar discussions. Electronic Journal of Foreign Language Teaching, *5*, 89-97.

Langer, I., Schulz von Thun, F., & Tausch, R. (1999). Sich verständlich ausdrücken [Expressing oneself comprehensively]. München, Germany: Ernst Reinhardt Verlag.

Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. Journal of Second Language Writing, *17*, 144-164.

Leki, I. (1991). The preferences of ESL students for error correction in college-level-writing classes. Foreign Language Annals, *24*, 203-218.

Leung, K., Su, S., & Morris, M. W. (2001). When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback. Human Relations, *54*, 1155-1187.

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. Journal of Computer Assisted Learning, *17*, 420-432.

Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. Active Learning in Higher Education, *7*, 51-62.

Lockhart, C., & Ng, P. (1993). How useful is peer response? Perspectives, *5*, 17-29.

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

Miedema, J. L. (2004). Fairness and the self. Unpublished doctoral dissertation, Leiden University, Leiden, The Netherlands. (Kurt Lewin Institute Dissertation Series, 2003-9)

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), Handbook of research on educational communications and technology (2nd ed., pp. 745-783). New York: Erlbaum.

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. Western Journal of Nursing Research, *24*, 815-829.

Narciss, S. (2006). Informatives tutorielles feedback [Informative tutorial feedback]. Münster, Germany: 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. Brünken (Eds.), Instructional design for multimedia learning (pp. 181-195). Münster, Germany: Waxmann.

Pezdek, K. (2001). Wiederentdeckte Erinnerungen oder falsche Erinnerungen? [Rediscovered memories or false memories] (J. Grabowski & E. Van der Meer, Trans.). In R. L. Atkinson, R. C. Atkinson, E. E. Smith, D. J. Bern, & S. Nolen-Hoeksema (Eds.), Hilgards Einführung in die Psychologie (p. 297). Heidelberg, Germany: Spektrum.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. Assessment & Evaluation in Higher Education, 33, 143-154.

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

Prose, A. (2006). Entwicklung und Evaluation computerbasierter Trainingsaufgaben für das wissenschaftlichen schreiben [Design and evaluation of computer-based training assignments for academic writing]. Unpublished doctoral dissertation, Technical University Dresden, Dresden, Germany.

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. Personality and Social Psychology Bulletin, 27, 1520-1532.

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

Slobin, D. I. (2001). Wie die Sprache das Denken lenken kann: linguistische Relativität und linguistischer Determinismus [How speech directs thought: Linguistic relativity and linguistic determinism] (J. Grabowski & E. Van der Meer, Trans.). In R. L. Atkinson, R. C. Atkinson, E. E. Smith, D. J. Bern, & S. Nolen-Hoeksema (Eds.), Hilgards Einführung in die Psychologie (p. 336). Heidelberg, Germany: Spektrum.

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

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

Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. Research in the Teaching of English, 31, 91-119.

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Needham Heights, MA: Allyn & Bacon.

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

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? Journal of Second Language Learning, 9, 147-170.

Van den Berg, I. (2003). Peer assessment in universitair onderwijs: Een onderzoek naar bruikbare ontwerpen [Peer assessment in higher education: A study on useful designs]. Unpublished doctoral dissertation, Utrecht University, Utrecht, the Netherlands.

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. Educational Research Review, 4, 41-51.

Van Gennip, N. A. E., Segers, M. S. R., Tillema, H. H. (this issue). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. Learning and Instruction.

Van Gog, T., Paas, F., Savenye, W., Robinson, R., Niemczyk, M., Atkinson, R., Johnson, T. E., O'Connor, D. L., Rikers, R. M. J. P., Ayres, P., Duley, A. R., Ward, P., & Hancock, P. A. (2008). Data collection and analysis (section on assessment of complex performance). 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. 763-800). Mahwah, NJ: Erlbaum.

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

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (this issue). Effective peer assessment processes: Research findings and future directions. Learning and Instruction.

Walker, A. (2001). British psychology students' perceptions of group-work and peer assessment. Psychology Learning and Teaching, 1, 28-36.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. Journal of Second Language Writing, 15, 179-200.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. Journal of Second Language Writing, 4, 209-222.

Table 1
Correlation between fairness, usefulness, acceptance, willingness to improve and affect ($N = 71$)

| | Fairness | Usefulness | Acceptance | Willingness | Affect |
|-------------|----------|----------------|------------|----------------|--------|
| Fairness | -- | | | | |
| Usefulness | .68* | --- | | | |
| Acceptance | .72* | .61* | -- | | |
| Willingness | .42* | .52* | .35* | -- | |
| Affect | .48* | .17, <u>ns</u> | .38* | .17, <u>ns</u> | -- |

* $p < .01$.

Table 2
Factor loadings of feedback perceptions items (four-factor oblimin principal component analysis)

| Scale | Items | I | II | III | IV | |
|----------------------------|---|-------------|-------------|-------|-------------|-------------|
| Fairness | I would be satisfied with this feedback | .665 | -.322 | .174 | .228 | |
| | I would consider this feedback fair | .479 | .334 | .145 | .321 | |
| | I would consider this feedback justified | .652 | .268 | .078 | .227 | |
| Usefulness | I would consider this feedback useful | .795 | .258 | -.080 | -.200 | |
| | I would consider this feedback helpful | .737 | .178 | .062 | -.129 | |
| | This feedback would provide me a lot of support | .682 | .128 | .201 | -.063 | |
| Acceptance | I would accept this feedback | .756 | -.071 | .007 | -.123 | |
| | I would dispute this feedback | .489 | -.006 | -.156 | .560 | |
| | I would reject this feedback | .569 | .034 | -.181 | .504 | |
| Willingness | I would be willing to improve my performance | .157 | .779 | -.122 | .067 | |
| | I would be willing to invest a lot of effort in my revision | .160 | .823 | .020 | -.036 | |
| | I would be willing to work on further text revision assignments | -.141 | .846 | .083 | .103 | |
| Affect | I would feel ... if I received this feedback on my revision | | | | | |
| | Positive | Satisfied | .242 | -.098 | .790 | .188 |
| | | Confident | -.027 | .042 | .928 | -.048 |
| | | Successful | -.007 | .022 | .929 | .011 |
| | Negative | Offended | -.202 | .019 | .057 | .848 |
| | | Angry | .027 | .096 | .064 | .811 |
| | | Frustrated | -.123 | .050 | .148 | .799 |
| Eigenvalue | | 6.44 | 1.58 | 2.78 | 1.97 | |
| % of variance explained | | 35.75 | 8.78 | 15.46 | 10.97 | |
| <i>Factor correlations</i> | | I | II | III | IV | |
| I | | | .282 | .150 | .225 | |
| II | | | | .017 | .138 | |
| III | | | | | .226 | |

Fairness, usefulness and acceptance items form the factor Perceived Adequacy of Feedback (I).

II = Willingness to Improve; III = Positive Affect; IV = Negative Affect.

Loadings above .400 are boldface.

Table 3

Mean and standard error for Perceived Adequacy of Feedback (PAF), willingness to improve (WI) and affect (AF) as a function of group

| | CGF-H (N = 18) | | CGF-L (N = 18) | | ESF-H (N = 18) | | ESF-L (N = 17) | |
|-----|----------------|-----------|----------------|-----------|----------------|-----------|----------------|-----------|
| | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> |
| PAF | 5.35 | 1.88 | 4.06 | 1.88 | 6.17 | 1.45 | 5.37 | 1.78 |
| WI | 6.58 | 2.36 | 5.69 | 2.31 | 6.51 | 1.96 | 7.22 | 1.71 |
| AF | 4.81 | 1.73 | 3.77 | 1.85 | 3.80 | 1.68 | 4.79 | 1.13 |

CGF-H = concise general feedback by a high competent peer; CGF-L = concise general feedback by a low competent peer; ESF-H = elaborated specific feedback by a high competent peer; ESF-L = elaborated specific feedback by a low competent peer.

Table 4

Mean and standard error for each group for knowledge of sub-criteria in pretest and posttest, revision performance by phase, time on task by phase, and revision performance by time on task by phase

| | CGF-H (N = 18) | | CGF-L (N = 18) | | ESF-H (N = 18) | | ESF-L (N = 17) | | Control (N = 18) | |
|-----------|--------------------------------------|-----------|----------------|-----------|----------------|-----------|----------------|-----------|------------------|-----------|
| | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> | <u>M</u> | <u>SD</u> |
| | Sub-criteria | | | | | | | | | |
| Pretest | 3.78 | 1.31 | 4.17 | 1.79 | 2.83 | 2.09 | 3.24 | 2.11 | 3.33 | 1.88 |
| Posttest | 3.83 | 3.11 | 4.59 | 2.94 | 4.14 | 2.53 | 4.72 | 1.91 | 4.22 | 2.65 |
| | Revision performance | | | | | | | | | |
| Pretest | 2.44 | 1.38 | 2.33 | 1.33 | 2.06 | 1.59 | 1.92 | 1.12 | 2.11 | 1.68 |
| Treatment | 5.61 | 2.57 | 5.94 | 3.73 | 3.67 | 2.40 | 4.53 | 2.10 | 6.22 | 3.10 |
| Posttest | 5.44 | 2.59 | 5.94 | 3.21 | 4.17 | 2.45 | 5.25 | 3.05 | 5.33 | 3.11 |
| | Time on task | | | | | | | | | |
| Pretest | 16.75 | 2.72 | 16.61 | 2.79 | 17.88 | 4.07 | 16.79 | 3.19 | 17.67 | 3.50 |
| Treatment | 19.03 | 3.31 | 19.55 | 6.20 | 20.72 | 6.08 | 22.67 | 4.90 | 18.45 | 5.64 |
| Posttest | 17.78 | 4.47 | 16.61 | 4.55 | 15.37 | 3.98 | 13.82 | 3.45 | 17.54 | 4.06 |
| | Revision performance by time on task | | | | | | | | | |
| Pretest | 0.14 | 0.06 | 0.14 | 0.07 | 0.11 | 0.08 | 0.11 | 0.07 | 0.12 | 0.09 |
| Treatment | 0.30 | 0.15 | 0.28 | 0.16 | 0.20 | 0.15 | 0.20 | 0.09 | 0.36 | 0.19 |
| Posttest | 0.30 | 0.13 | 0.36 | 0.20 | 0.26 | 0.14 | 0.38 | 0.20 | 0.31 | 0.18 |

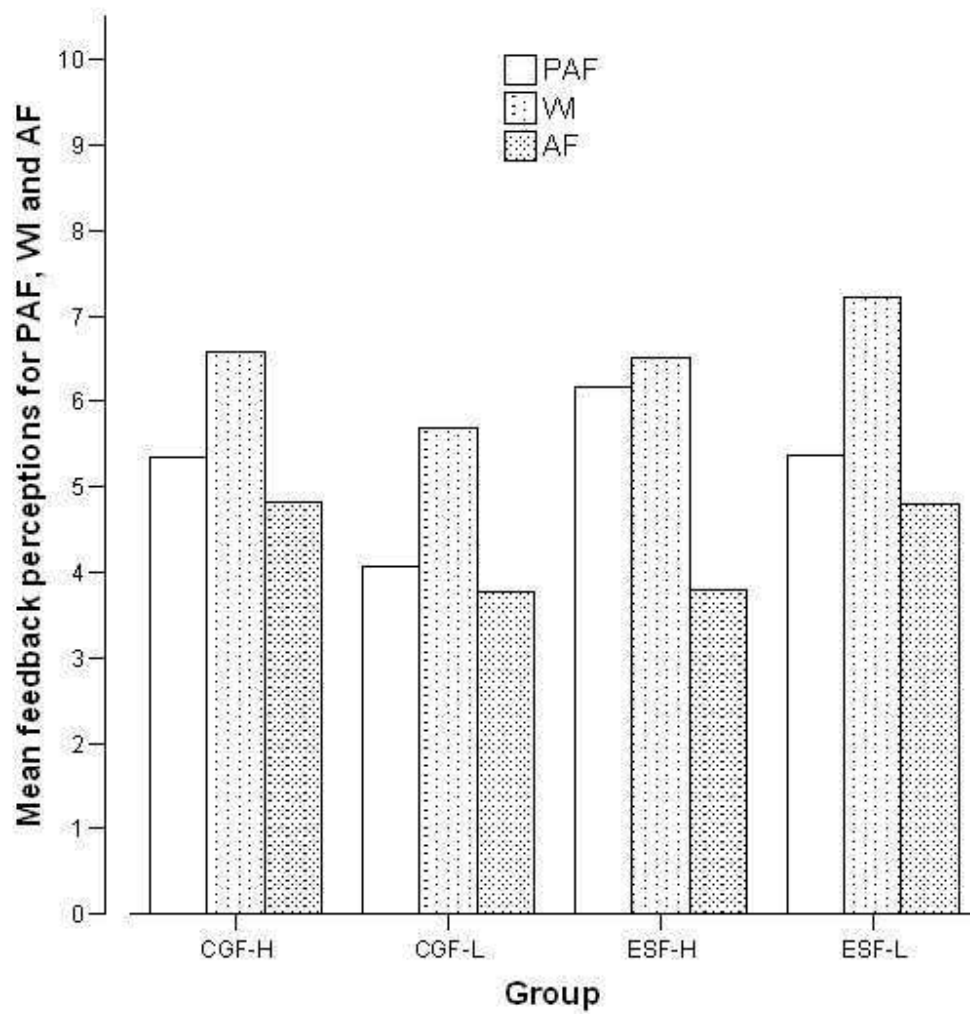


Fig. 1. Mean feedback perceptions for PAF, WI and AF as a function of group.

Figure 2

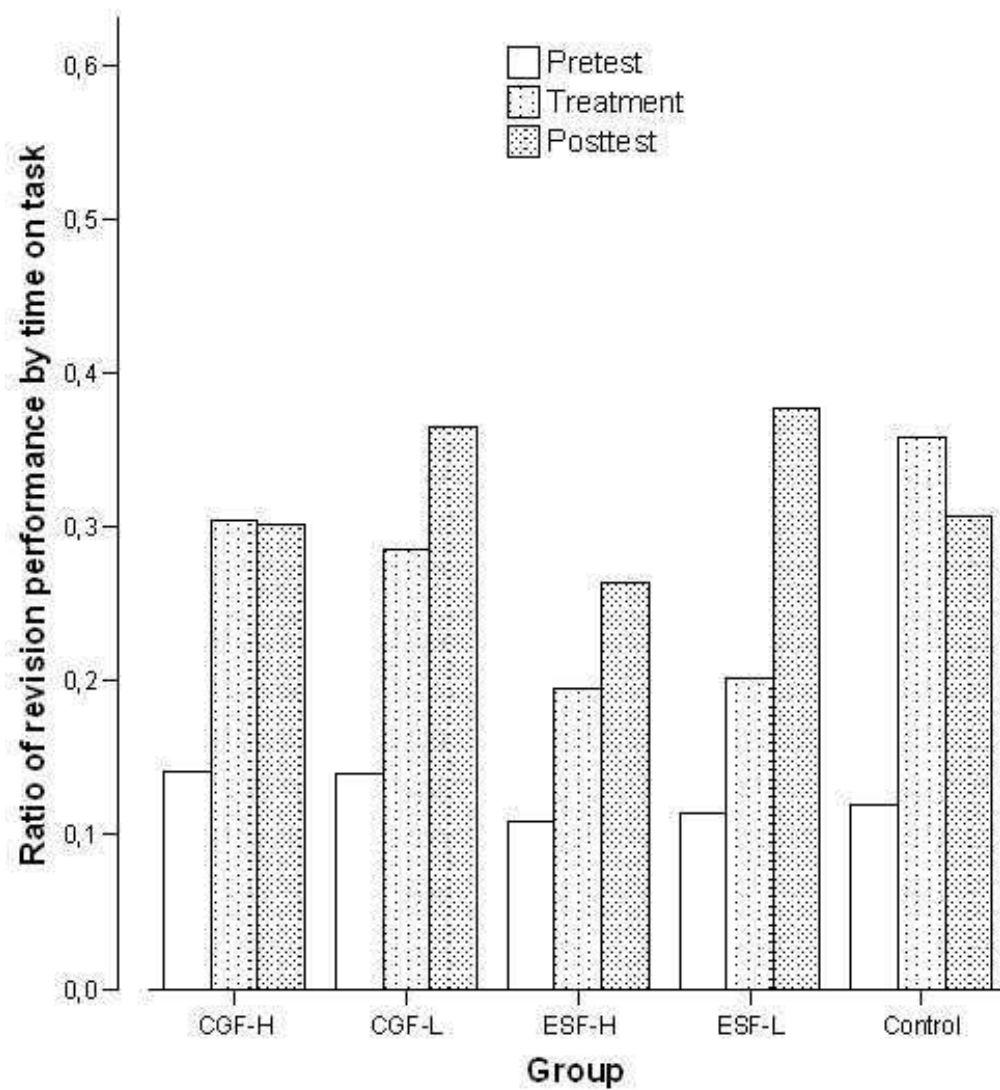


Fig. 2. Revision performance by time on task as a function of phase and group.