Leveraging the potential of peer feedback in an academic writing activity through sense-making support

Astrid Wichmann[1] · Alexandra Funk[1] ·
Nikol Rummel[1]

Abstract The act of revising is an important aspect of academic writing. Although revision is crucial for eliminating writing errors and producing high-quality texts, research on writing expertise shows that novices rarely engage in revision activities. Providing information on writing errors by means of peer feedback has become a popular method in writing instruction. However, despite its popularity, students have difficulties in leveraging the potential of peer feedback: feedback uptake is low and students engage in little revision. Instructional support might help learners to make sense of peer feedback and to reflect on the provided information more deeply. The present study investigated the effect of sense-making support on feedback uptake as well as on revision skills, in particular problem detection and problem correction. In an experimental study, 73 university students were randomly assigned to conditions with or without sense-making support. The results indicate that feedback uptake improved concerning two out of three variables: students in the condition with sense-making support made fewer new errors and rejected more incorrect feedback comments. Students' revision skills only improved with regard to problem detection. Overall, we were able to show that peer feedback alone might not be sufficient to make successful changes in the text and improve revision skills. Sense-making support proved to be effective to some extent and partially helped to maximize the benefits of peer feedback.

Keywords Peer assessment · Feedback uptake · Academic writing

Introduction

When we receive feedback from others on composed documents or paragraphs of text, we usually appreciate it. Yet, there is a tendency to ignore parts of the feedback, especially when it suggests

✉ Astrid Wichmann
   astrid.wichmann@rub.de

[1] Educational Psychology, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, NRW, Germany

revisions, and some feedback therefore remains unused. Instructional support that helps the writer to make sense of feedback may assist the writer to use feedback, which would otherwise have been neglected, for text revision. In this study, we observed the reception of peer feedback by students, in other words, writers who often have very little writing practice. First, we argue that the provision of feedback does not guarantee its implementation by the writer. Second, we claim that inexperienced writers engage in little revision, even though finding errors and changing the text accordingly are important skills in academic writing. We suggest that peer feedback may assist writers by detecting problems or errors. However, in order to improve revision skills, it might not be sufficient to solely provide peer feedback. Against this background, we experimentally investigated whether sense-making support during feedback provision helps students to increase feedback uptake and to improve revision skills, specifically problem detection skills and problem correction skills.

Peer feedback in academic writing

In this paper, we discuss the benefits of peer feedback as it relates to feedback uptake and revision skills in an academic writing activity. Peer feedback typically includes the following overt activities: first, the assessee creates a product (task performance). Second, the assessor provides feedback (feedback provision). Third, the assessee needs to make sense and form a coherent picture of the feedback (feedback reception). Lastly, the assessee revises his/her own product based on the assessor's feedback (revision) (Kollar and Fischer 2010). Peer feedback has become a popular method for learning in university settings. Its use in academic courses may help students to improve their writing and to gain knowledge of the subject matter (Falchikov 1986; Roscoe and Chi 2007). Feedback from peers can be just as effective and sometimes even better for gaining writing skills compared to feedback from a teacher (e.g., Topping 1998). In peer feedback, people of similar or equal academic status exchange feedback (Goldin et al. 2012). Within the context of writing instruction, it can be implemented in various ways. Typically, an assessee writes a text and then receives comments from an assessor. The assessor's feedback highlights errors in the text, which help the assessee with problem detection. Furthermore, the assessor may include suggestions on how to make changes in the text during error correction. After feedback reception, the assessee needs to decide how to use the feedback. The assessee may reject the feedback because he/she perceives the problem as too trivial or too difficult to fix, resulting in no revision. Alternatively, the assessee may perceive the feedback as relevant, and corrects the problem, thereby revising the text (Hayes 2004). Based on the assumption that students' time management is more flexible than that of teachers, peer feedback can be received "just in time," in contrast to expert feedback (Falchikov and Goldfinch 2000). Moreover, as there are more students than teachers, it is possible to provide more peer feedback than expert feedback. Peer feedback not only complements instructors' feedback (e.g., Hammer et al. 2010; Zariski 1996), but is increasingly becoming an independent form of feedback in itself. In the research on peer feedback, a shift can be discerned from the mere focus on aspects of reliability and validity, towards viewing peer feedback as a social process. Providing and receiving peer feedback can be inherently understood as a collaborative activity, which holds rich learning opportunities for both the assessor and assessee (Falchikov and Goldfinch 2000).

Assessees' difficulties in using peer feedback during writing

We claim that the provision of peer feedback does not guarantee that the recipient will benefit from it. Although peer feedback holds rich possibilities for students to improve their learning, students have problems in leveraging its potential. One reason for this relates to the way in which students make use of peer feedback. Often, they handle it ineffectively, either because they are reluctant to use it or because they reject the feedback upfront and ignore the contained information (Boero and Novarese 2012). Such difficulties in using feedback effectively have been described as failure of feedback uptake (Van der Pol et al. 2008). Feedback uptake is the ability to use feedback for the purpose of improving the text. In other words, through feedback uptake, students consider feedback, re-evaluate the text based on feedback, and possibly make changes which improve the text. In the case of peer feedback, feedback uptake is an important aspect that requires greater research attention. Inexperienced students struggle to consider peer feedback and to use it effectively to improve their texts. Yet, feedback uptake is particularly important for inexperienced writers, as they might be unable to detect text problems on their own.

There are several potential reasons why students struggle with feedback uptake (e.g., Van Gennip et al. 2010; Nelson and Schunn 2009). One such reason refers to a lack of reflection on received feedback: assessees do not sufficiently engage in reflection on their own. However, research on self-regulation suggests that reflection is crucial for the process of acquiring knowledge and new skills (Zimmerman 1989), and it is particularly important for students assuming the role of the assessee. We argue that by reflecting, assessees maximize the benefits from peer feedback. Reflection includes three core processes: (1) planning, (2) monitoring, and (3) evaluation (Schraw 1998). Planning encompasses understanding one's own knowledge of feedback, e.g., knowing how peer feedback (assessors' intentions) relates to the meaning conveyed in the text (assessees' intentions). Monitoring involves keeping track of which feedback the assessee agrees with and which feedback he/she rejects. Evaluation comprises, for example, which feedback the assessee rejects and which feedback will be used to make changes in the text. Another problem lies in the limited knowledge regarding how to handle the information delivered through feedback. Students may not know how to make changes in the text based on peer feedback because they lack a model of how to work through text problems systematically. Expert writers have the ability to represent detected writing problems by conducting a means-ends analysis. This gives them a better understanding than novice writers regarding what actions need to be taken to correct problems and to successfully revise text (Hayes 2004; Newell and Simon 1972). In sum, the provision of peer feedback might not be sufficient to ensure feedback uptake due to insufficient reflection and a lacking model of how to tackle text problems systematically.

Students' difficulties in problem detection and problem correction

Furthermore, we argue that students engage in little revision as they lack the necessary skills to detect problems on their own. Particularly at the beginning of an academic career, students struggle with writing. While it might be relatively easy to develop a first draft of a text, revising the draft has been shown to be very difficult for students. The importance of revision has been discussed in the area of writing research (MacArthur et al. 2006; Alamargot and Chanquoy 2001). Engaging in revision during writing may affect the text quality as well as learning (Flower et al. 1986; Hayes 2004). Revision can be defined as a process that leads to changes in which writers engage during writing (Fitzgerald 1987). As writers revise a text,

they re-organize ideas and integrate new ideas with existing ones, generating a coherent line of argument. Revision can also be viewed as a skill consisting of subskills, including problem detection and problem correction (Hayes et al. 1987). The detection of a problem or an error presupposes that the writer identifies a gap between the intended text and the text that has been produced (Fitzgerald 1987). To detect and diagnose problems correctly, some understanding of the nature of textual errors and writing criteria is required (Hayes et al. 1987). Problem detection skills entail the general capacity to recognize that a sentence or text section is erroneous. However, effective problem correction depends on diagnosing the error (Hayes and Flower 1986). Correcting a problem involves three steps: deciding what needs to be changed, how the desired changes are to be made, and how the changes are to be instantiated in the text (Hayes 2004).

Students face several problems during revision. As inexperienced writers, they spend too little time revising in an earlier stage of the draft, and do so too superficially (Allal and Chanquoy 2004; Hayes et al. 1987; Proske et al. 2010). For successful revision, in a first step, a writer needs to detect a problem, which can be a difficult undertaking. Novice writers detect fewer problems in texts than do experts: a study by Hayes and colleagues (Hayes et al. 1987) showed that experts detected 58% of planted problems, while novices detected only 36%. Difficulties in problem detection arise from unclear goal representation. An inexperienced writer might have a fuzzy idea of what a text should convey. A further difficulty lies in writing a text from the reader's perspective. Inexperienced novice writers tend to forget the audience for which they are writing. Furthermore, compared to experts, inexperienced writers have a less-developed set of writing criteria in mind (Graham & Harris, 2007). However, as mentioned above, knowledge about writing criteria is crucial for detecting errors. Knowledge about criteria for good writing, such as typical writing errors, enables writers to identify errors and categorize error types. A second step during revision is to correct the problem or error, which can still be challenging even once the error has been detected. In sum, students lack skills regarding both problem detection and problem correction. Help from peers, who point out errors, might foster writers' awareness of problems in the text (Hayes 2004). Nevertheless, as described above, peer feedback alone might not be sufficient to improve problem detection and problem correction skills.

Helping the assessee to leverage the potential of peer feedback during writing

We argued above that improving feedback uptake and revision skills might require not only peer feedback but also instructional support. Combining peer feedback with instructional support should be helpful for tackling the aforementioned problems. First, instructional support should focus on reflection (or lack thereof). Similarly to learning protocols (Berthold et al. 2007), assessees should write down their reflections on peer feedback. Support should instruct assessees to think about whether the feedback was understood (planning), to determine whether there is a gap between the assessor's intentions and the assessee's intentions (monitoring), and to judge whether feedback is considered to be usable (evaluation) and how it will be used to improve the text.

Second, support should help assessees to work through text problems more systematically. Just as expert writers do, students might then represent problems as part of a means-ends table. This should help students to represent the relation between detected text problems and actions that need to be taken to correct a problem. Third, assessees should be supported to reflect more deeply on feedback. Analogously to learning protocols (Berthold et al. 2007), assessees should write down their reflections on peer feedback. In sum, support should instruct assessees to

think about whether the feedback was understood (planning), to determine whether there is a gap between the assessor's intentions and the assessee's intentions (monitoring), and to judge whether feedback is considered to be usable (evaluation) and how it will be used to improve the text. By both reflecting on and tackling text problems systematically, we aim to help assessees to make sense of peer feedback.

Research questions and hypotheses

In this paper, we aim to explore ways to support the assessee during peer feedback. There is substantial evidence that assuming the role of the assessor by providing feedback and assessing products created by peers leads to learning gains (Topping 2003). However, there is little empirical evidence that assuming the role of the assessee, that is, the feedback recipient, leads to learning gains as well (Van der Pol et al. 2008; Cho and MacArthur 2010; Kluger and DeNisi 1996).

Previous studies in the context of peer feedback research did not systematically tease apart the role of the assessor and the assessee (Li et al. 2010; Van Zundert et al. 2010). Moreover, the robustness of study findings was limited, as attention was not paid to reducing the variability of extraneous variables and the chance of confounding effects. In particular, research did not systematically control for the type of peer feedback (Nelson and Schunn 2009). Finally, peer feedback research rarely differentiates between performance and learning. Our study addressed these shortcomings as follows: to tease apart the role of assessor and assessee, we conducted an experiment in which all participants served as assessees. Furthermore, to reduce the variability of extraneous factors, we controlled for the type of feedback by choosing trained tutors (acting as peers) instead of real peers for peer feedback provision. Finally, we differentiated between performance and learning by conducting a content analysis assessing feedback uptake and measuring revision skills in a pre-posttest design.

Overall, we explored whether instructional support, in the form of sense-making support, helps assessees to maximize the benefits from peer feedback. Based on our assumptions presented above, we expected that sense-making support would facilitate feedback uptake (hypothesis 2). Since feedback uptake was conceptualized as changes that assessees make based on received peer feedback, we assumed that students would incorporate more feedback-based revisions in the text. Furthermore, we assumed that sense-making support would improve assessees' revision skills (hypothesis 1). As revision skills were subdivided into problem detection and problem correction, we hypothesized that both subskills would improve as a result of sense-making support.

Method

Participants and study design

Participants were from a German University studying towards a degree in Education. Participation was a mandatory part of their regular class activities; however, no grades were given for the writing activity. The study was conducted in seven entry-level university courses, as it was assumed that writing proficiency at that level would be low, and that experience of receiving peer feedback would be limited. Data were gathered in two cohorts. As there were no significant differences between the cohorts with respect to demographic variables, performance scores on all tests, or feedback uptake scores, the two cohorts were merged for the main analysis. Altogether, 125 students participated, of whom $n = 52$ were excluded from the main

analysis for the following reasons: non-native speakers ($n = 19$) were excluded because writing posed a much greater challenge for these students and $n = 33$ students were excluded due to incomplete data (e.g., pretest data were missing). Thus, data from $n = 73$ students (12 males, 61 females) were used for the main analysis. Students were on average in their $M = 2.26$ (SD = 1.65) semester and their mean age was $M = 23.37$ (SD = 3.37) years. Each participant was randomly assigned to one of two conditions. Altogether, there were 34 participants in the "sense-making support (SMS+)" condition and 39 in the "no sense-making support (SMS−)" condition.

Learning activity

Students participated in an online writing activity. The activity consisted of the phases "Draft"—"Feedback Reception"—"Revision." The task was to write an essay of 550–650 words on the question: "How can a mother optimally support the identity formation of her child?" As background information, students received an excerpt of a text by Erikson(1959) on identity formation. The writing task was accompanied by peer feedback.

Description and implementation of sense-making support (independent variable)

Sense-making support was delivered during the phase "PeerFeedback Reception" in an MS Word document. The document listed each feedback comment in a table. Students in the SMS+ condition were instructed to attend to each feedback comment, completing the table using check boxes and open responses.

Sense-making support aimed at encouraging participants to reflect on their understanding of the feedback comments in terms of planning, monitoring, and evaluating feedback. In the first column of the table (column 1), participants were asked to list each feedback comment using the copy/paste function. They were then asked to judge each comment in the list with regard to understanding (column 2), agreement (column 3), plans to use feedback (column 4), plans on how to use feedback (column 5) as well as monitoring (column 6) and relevance (column 7) of feedback (Fig. 1).

Procedure

The online writing activity was conducted during a period of 2 weeks (see Fig. 2). The learning platform Moodle was customized for the online writing activity (Moodle 2013). We followed the structure of the peer feedback phases identified by Kollar and Fischer (2010), focusing on the role of the assessee.

During draft writing, students received instruction and background information on how to write the essay (see Table 1). They were asked to upload the essay as an MS word document.

1	2		3		4		5	6	7
Copy your received comments into this column.	I understand the comment.		I agree with the comment.		I am going to use this comment		I will improve my essay by doing the following:	Done	Mark the three most important comments with an X.
	Yes	No	Yes	No	Yes	No		✓	
Comment 1	☐	☐	☐	☐	☐	☐		☐	
…	☐	☐	☐	☐	☐	☐		☐	

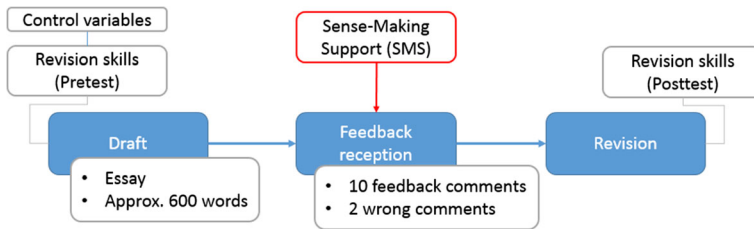**Fig. 1** Sense-making support: table listing each feedback comment

Fig. 2 Online-writing activity with instruments and peer feedback phases

The students did not provide feedback themselves; all students received feedback from tutors (acting as peers). Tutors were trained by first receiving a manual that specified all writing criteria as well as typical errors. Additionally, they practiced by providing feedback on student texts. A researcher then reviewed and corrected feedback until the tutor's feedback comments corresponded to the guidelines in the manual. Additionally, all feedback comments were checked by a second tutor before feedback provision. During feedback reception, participants received peer feedback. During revision, the essays, including feedback comments, were made available and participants revised their essays. Afterwards, participants uploaded their revised documents to Moodle. Participants were guided through each phase on a step-by-step basis. To start a new phase, the participants needed to have completed the previous one.

As described above, all participants assumed the role of the assessee. Participants were informed that the feedback was given by a peer. However, in order to control for the variance of feedback quantity and quality, feedback was prepared by trained tutors, as described above. Peer feedback included 12 comments for each participant. Each comment referred to one error in the text and included a standardized description of the error and a suggestion on how to revise it (see Fig. 3).

Comments were given based on common writing errors types (e.g., Esselborn-Krumbiegel 2010; Kornmeier 2008; Kruse 2010; Krämer 1999). We provided comments on the five most typical error types found in writings by our freshman students, including Sequence/Logic of Argument, Transition Words, Nested Sentences, Direct/Clear Reference, and Filler Words.

The comment structure was as follows: "This sentence is hard to read. Rephrase it in order to make it more readable." Each student received two correct comments per error type, plus two incorrect comments, which makes 12 comments overall. The incorrect comments were included in order to mimic real peer feedback, which is prone to be erroneous. Feedback was delivered in the drafts that the students had earlier uploaded using the commenting function of MS Word. For each writing error, we highlighted the relevant section of the text. For filler words, we highlighted the word itself; for missing transition words, we highlighted the last word and the first word of the adjacent sentence. For the remaining error types, we highlighted the whole sentence.

Feedback was provided following a rigorous procedure. First, tutors read an essay in order to get an impression of its intended statement and logical structure. Next, the essay was re-read

Table 1 Online writing activity including times for peer feedback phases

Peer assessment phases	Time on task (min)
Draft	120
Feedback reception	10
Revision	80

Vertrauen. Dieses Vertrauen bietet die Grundlage für den Beginn der Identitätsbildung. Besonders wichtig ist hier, dass die Mutter eine konstante Beziehung zum Kind aufbaut. Um seine innere Welt zu konstruieren, und somit Identität zu erlangen, benötigt der Säugling kontinuierlich wiederkehrende Situationen mit den gleichen Personen in seinem Umfeld. Nur so können die Erlebnisse innerlich aufgenommen werden.

Filler word | Transition word | Logical sequencing

Comment
Logic. This sentence does not logically follow the previous sentence. Make sure the logic is clear.

Comment
Connect sentences. The sentences are logically linked, yet a connector is missing. Try to connect the sentences, emphasizing the link.

Comment
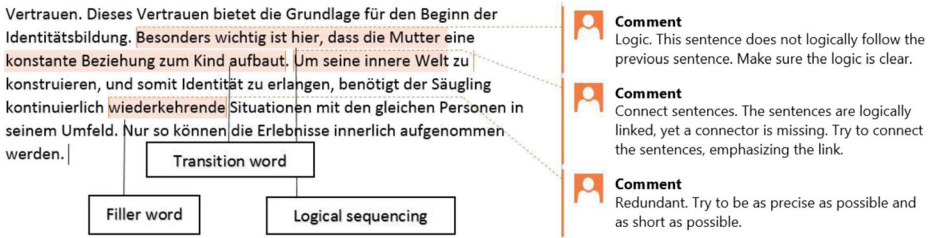Redundant. Try to be as precise as possible and as short as possible.

Fig. 3 Examples of highlighted errors (comments translated)

and commented on by focusing on one writing error at a time. Tutors read each essay at least six times. A standardized comment was given for each error type. Of all errors in the text, the ten most distinct ones were commented on. In addition, two erroneous comments were added, mimicking real peer feedback under the assumption that peer feedback is not always correct.

Measures and instruments

Control measures and manipulation check We controlled for uneven distribution in the conditions taking into account demographic information and experience with and interest in using computers. To discern whether sense-making support was used as intended, we analyzed participants' attendance to the sense-making support table.

Feedback uptake Feedback uptake was assessed by measuring feedback-based changes in the text. We distinguished between "successful change (yes/no)," "New error change (yes/no)," and "incorrect comments change (yes/no)." All texts were assessed by one coder and 25% of the texts were coded by a second coder. For all analyses, coders were unaware of the treatment conditions. Coders' percentage of agreement was between 87 and 91% (see Table 2).

Revision skills (pre-posttest) Revision skills were assessed using two counterbalanced pre- and posttest versions A and B (Table 3). Items in both versions were the same with respect to structure but differed concerning content. Versions A and B were randomly administered to both conditions. The pre- and posttests assessed two distinct skills related to academic writing: problem detection and problem correction. Problem detection was assessed with an erroneous text. As described, problem detection skills include recognizing erroneous text and diagnosing the problem. We operationalized these steps by asking participants to highlight and label text passages they perceived as erroneous. The number of correctly highlighted errors informed us about whether students generally recognized errors. The number of labels added showed us

Table 2 Description and reliability of feedback uptake variables

Feedback uptake variables	Description	Agreement %	Kappa
Successful change	Any change that resulted in an improvement of the text	87.3	.66
New error change	Any change that erroneously created a new mistake instead of correcting one	91.3	.47
Incorrect comment change	Any change that was made based on an incorrect feedback comment	90.1	.82

Table 3 Reliability of revision skills: problem detection	Cronbach's α
Pretest version A	.61
Pretest version B	.81
Posttest version A	.67
Posttest version B	.78

whether students could describe the nature of the error. The maximum score was 20 points, consisting of 10 points for correct highlighting and 10 points for correct labeling.

For problem correction, participants had to correct errors in text sections. Errors were related to the writing errors described earlier. The A and B test version included items concerning Sequence/Logic of Argument, Transition Words, Nested Sentences, Direct/Clear Reference, and Filler Words. Points differed for each item (Table 4). A maximum score of 22 points could be achieved.

Results

The reported results can vary with respect to number of participants, because not all of the 73 participants completed all relevant stages. All measures where checked for homogeneity of variances and normality. In the case of violation of the normality assumption, non-parametric tests were conducted.

Control measures and manipulation check

We controlled for uneven distribution in the conditions, taking into account demographic information and experience with and interest in using computers. There were no substantial differences between participants of the two conditions regarding prior experience with computers ($F(1, 71) = .02$, $p = .88$) and interest in computers ($F(1, 71) = .46$, $p = .50$). Moreover, we conducted a manipulation check in order to verify that the participants used sense-making support as intended and in order to exclude participants who used sense-making support in a fragmentary or incorrect manner. Students in the SMS+ condition received a table which helped them to attend to and reflect on each feedback comment, and were asked to complete the table using check boxes and open responses. We analyzed participants' adherence by checking how they completed the table in terms of attending to each feedback comment (Fig. 1). We were interested in whether participants ($N = 34$) adhered to each of the columns of the table, particularly columns 2 ("I understand the comment"), 3 ("I agree with the comment"), and 4 ("I am going to use the comment"). Furthermore, we checked whether

Table 4 Scores for problem correction	Criteria	Score
	Sequence/Logic of Argument	5
	Filler Words	6
	Transition Words	5
	Direct/Clear Reference	2
	Nested Sentences	4
	Total	22

participants followed up on what they reported in the table by using the feedback comments to revise their texts accordingly. In other words, the aim was to establish whether those who reported using feedback comments actually did so when revising their texts based on the feedback. Thus, the number of comments that was reportedly used was compared to the number of comments that was actually used for revision. The same applied for participants' handling of erroneous feedback comments: we investigated the extent to which participants reported using erroneous comments and whether they followed up on them when revising their texts.

All participants completed the table, and did so in the intended manner. Out of 10 feedback comments, participants reported that they understood 85.9%, agreed with 61.8%, and used 70.3% (see Table 5). Participants followed up on most of the feedback comments they intended to use: approximately 80% of the comments that participants reported using were actually used for revision. Participants were less inclined to use erroneous feedback comments, and did so to a much lesser extent compared to correct feedback comments. Of the two additional erroneous comments, participants reported that they understood 82.5%, yet only agreed with 41%, and only used 47%. Moreover, even fewer students followed up on these comments for revision. Out of all of the erroneous comments which students reported using, they actually used only 42.5%. Although it appears that students did use erroneous comments to some extent, the percentage of erroneous comments used was much lower than that of correct comments used. We found some discrepancy between what students reported in the table and their actual use of comments for revision. Over 60% (65.9%) reported using a feedback comment and indeed subsequently did so. The percentage was much lower for erroneous comments: only 35% reported using the comment and then actually did so. Based on these results, it appears that participants used sense-making support as intended, that is, for the purpose of reflection and tackling problems more systematically.

Effect of sense-making support on feedback uptake (content analysis)

The first goal of the study was to examine the effect of sense-making support on performance. A one-way multivariate analysis of variance (MANOVA) was conducted to determine the effects of sense-making support on feedback uptake with the variables successful change (SC), new error change (NC), and incorrect comment change (IC). The results showed a significant effect, Wilks' $\lambda = .85$, $F(1, 71) = 3.99$, $p < .01$, $\eta2 = .15$. Separate ANOVAs for the feedback uptake variables were then conducted in accordance with the corresponding MANOVA. The results showed that the average score for successful change, $F(1, 71) = 1.81$, $p < .18$, $\eta2 = .03$ did not differ between conditions. However, the feedback uptake variables, new error change,

Table 5 Means and standard deviations for participants' attendance to sense-making support table

	Feedback comments (10)			Additional erroneous feedback comments (2)		
	M	(SD)	%	M	(SD)	%
Column 2: "I understand the comment"	8.59	(1.84)	85.9	1.65	(.60)	82.5
Column 3: "I agree with the comment"	6.18	(8.57)	61.8	.82	(.80)	41.0
Column 4: "I am going to use the comment"	7.03	(5.85)	70.3	.94	(.78)	47.0
Actual use of feedback comments for revision	8.0	(1.65)	80.0	.85	(.70)	42.5

$F(1, 71) = 6.58$, $p < .01$, $\eta 2 = .09$ and erroneous comment change, $F(1, 71) = 4.14$, $p < .05$, $\eta 2 = .06$ were lower for students in the SMS+ condition than for students in the SMS− condition (see Table 6).

The changes reported above refer to feedback-based changes. Apart from feedback-based changes, students engaged in very little revision. Therefore, we do not report on results on revision other than feedback-based changes.

Effect of sense-making support on revision skills (pre-posttest)

The second goal of the study was to examine effects of sense-making support on students' learning, with a particular focus on their revision skills. The test comprised two parts: problem detection skills and problem correction skills.

Problem detection skills For problem detection, we distinguished between highlighted text passages pointing to an error and labeled errors including the error type. As we were interested not only in whether students engaged in highlighting and labeling activities in general but also in whether they did so correctly or incorrectly, we further distinguished between correctly and incorrectly highlighted text passages as well as correctly and incorrectly labeled errors. Tables 7 and 8 show the results for pretest and posttest, respectively. We included percentages for incorrectly highlighted and incorrectly labeled errors because the main unit for calculation differs between the two. The inclusion of percentages allows a direct comparison of both error types. The low number of participants resulted from a combination of corrupted files and participants who did not take part in the last phase of the online writing activity in which the posttest was conducted.

a) The number of highlighted text passages was $M = 8.45$ (SD = 3.91) for the pretest and $M = 8.22$ (SD = 3.7) for the posttest. An analysis of covariance with repeated measures, controlling for test version, did not reveal differences between pre- and posttest or between conditions ($F(1, 65) = .04$, $p = .83$; $F(1, 65) = .73$, $p = .40$).
b) The number of correctly highlighted text passages was $M = 4.77$ (SD = 1.97) for the pretest and $M = 4.91$ (SD = 2.01) for the posttest. An analysis of covariance with repeated measures, controlling for test version, revealed differences between pre- and posttest ($F(1, 65) = 8.57$, $p = .01$). Furthermore, the results showed a trend towards an interaction effect between test time and condition ($F(1, 65) = 3.53$, $p = .07$). Students in the SMS+ condition improved from pre- to posttest, whereas students in the SMS− condition barely improved.

Table 6 Means and standard deviations for feedback uptake

Revision types	SMS+ (N = 34)			SMS− (N = 39)		
	M	(SD)	Percentage of uptake	M	(SD)	Percentage of uptake
Successful change (SC)	5.06	(1.67)	50.6	5.59	(1.7)	55.9
New error change (NC)	1.03	(1.09)	10.3	1.72	(1.19)	17.2
Incorrect comment change (IC)	.85	(.70)	42.5	1.21	(.77)	60.5

Table 7 Means and standard deviations for problem detection in the pretest

Condition	Highlighted		Correctly highlighted		Incorrectly highlighted		Incorrectly highlighted (in %)		Correctly labeled		Incorrectly labeled (in %)	
	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)
With SMS	34	7.79 (4.54)	34	4.24 (2.15)	34	3.79 (2.79)	34	38.14 (23.87)	34	3.53 (1.99)	34	18.54 (36.15)
No SMS	39	9.03 (3.22)	39	5.23 (1.69)	39	3.56 (3.31)	39	38.26 (19.9)	39	4.31 (1.56)	39	14.78 (15.92)
All participants	73	8.45 (3.91)	73	4.77 (1.97)	73	3.68 (3.03)	73	38.21 (21.68)	73	3.95 (1.80)	73	16.53 (27.14)

Table 8 Means and standard deviations for problem detection in the posttest

Condition	Highlighted		Correctly highlighted		Incorrectly highlighted		Incorrectly highlighted (in %)		Correctly labeled		Incorrectly labeled (in %)	
	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)
With SMS	30	7.8 (4.32)	30	4.73 (2.16)	30	3.07 (3.27)	30	30.91 (26.21)	30	4.07 (1.95)	30	15.42 (23.78)
No SMS	38	8.55 (3.14)	38	5.05 (1.90)	38	3.50 (2.22)	38	38.17 (19.1)	38	3.79 (1.56)	38	20.21 (16.33)
All participants	68	8.22 (3.7)	68	4.91 (2.01)	68		68	34.79 (22.82)	68	3.91 (1.73)	68	17.98 (20.13)

c)   The number of incorrectly highlighted text passages was $M = 3.68$ (SD = 3.03) for the pretest and $M = 3.31$ (SD = 2.17) for the posttest. There was no significant difference between pretest and posttest ($\chi^2(1) = 1.143$, $p = .285$).

d)   The percentage of incorrectly highlighted text passages did not differ between pretest $M = 38\%$ (SD = 21.68) and posttest $M = 34.79\%$ (SD = 22.82).

e)   The number of correctly labeled text passages was $M = 3.95$ (SD = 1.80) for the pretest and $M = 3.91$ (SD = 1.73) for the posttest. Students in the SMS+ condition labeled 86.04% correctly and students in the SMS− condition labeled 75.05% correctly. An ANOVA with repeated measures showed an interaction effect between test time and condition ($F(1, 65) = 6.70$, $p = .01$). Students in the SMS+ condition improved from pretest to posttest, while students in the SMS− condition deteriorated from pretest to posttest.

f)   The percentage of incorrectly labeled text passages was $M = 16.53$ (SD = 27.14) for the pretest and $M = 17.98$ (SD = 20.13) for the posttest. There was no significant difference between pretest and posttest ($\chi^2(1) = 2.667$, $p = .102$) (see Figs. 4 and 5).

Our assumption was that students in the SMS+ condition would detect more problems due to increased possibilities to reflect on feedback. Indeed, we found a trend for improvement in the SMS+ condition in terms of correctly highlighting text passages, while no such improvement was found in the SMS− condition. Furthermore, students in the SMS+ condition improved in terms of correctly labeling errors, while students in the SMS− condition did not.

Problem correction skills The second part of the revision skills test assessed problem correction. Out of 22 possible points, students achieved $M = 75.06$ (SD = 14.59) percent in the pretest and $M = 76.73$ (SD = 9.70) in the posttest (Table 9). A univariate analysis with repeated measures did not reveal significant differences for either time (pre-posttest differences) or condition ($F(1, 47) = .85$, $p = .36$; $F(1, 47) = .09$, $p = .76$). In other words, problem correction skills were similar in the pre- and posttest.
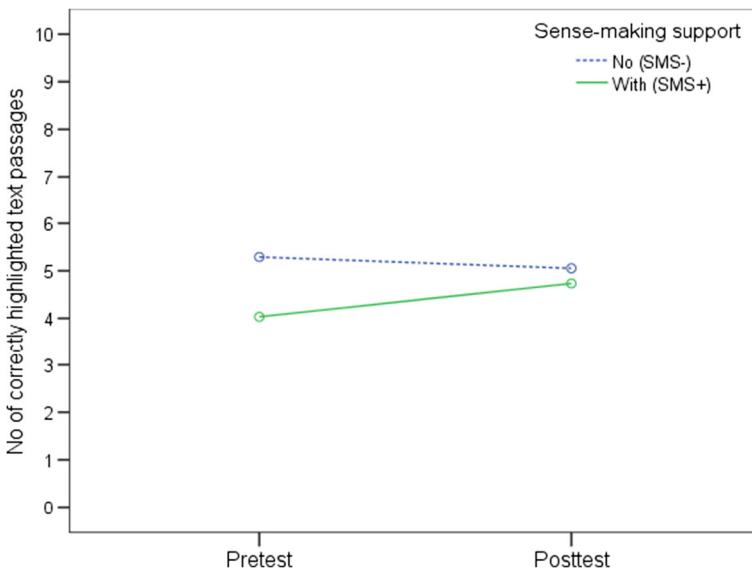

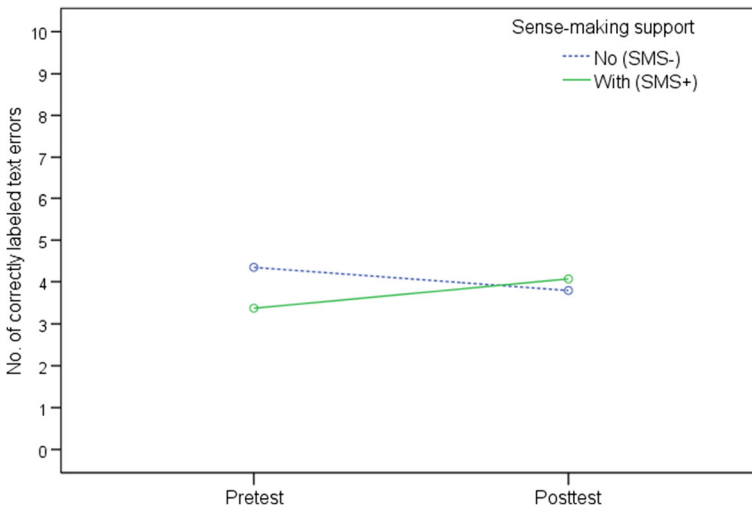Fig. 4   Number of correctly highlighted text passages

Fig. 5 Number of correctly labeled errors

We initially assumed that use of the table listing each feedback comment would enable students in the SMS+ condition to better reflect on feedback, and that they would therefore correct more problems in the posttest than students in the SMS− condition. This assumption was not confirmed by our findings.

Conclusions

In this study, we explored instructional support that aimed to help the assessee in the context of peer feedback. In particular, we looked at whether students used sense-making support for the intended purpose of reflection and tackling problems systematically. Sense-making support consisted of a table listing each feedback comment to help the assessee to plan, monitor, and evaluate his/her understanding of feedback. We focused on feedback uptake (performance) and revision skills (learning).

First, we wished to investigate whether sense-making support affected feedback uptake. Our results indicated that this was the case to some extent: students receiving sense-making support made fewer new errors. The treatment check analysis revealed that over 80% of the feedback comments were used to make successful changes in the text. We can conclude that students indeed used the table listing each feedback comment as a way to conduct a means-ends analysis. Students in the SMS+ condition seemed to relate mentioned errors to

Table 9 Means and standard deviations for problem correction

Condition	Pretest (%)		Posttest (%)	
	n	M (SD)	n	M (SD)
SMS+	26	75.32 (14.89)	24	77.24 (9.11)
SMS−	32	74.85 (14.58)	36	76.4 (10.19)
Total	58	75.06 (14.59)	60	76.73 (9.70)

actions in order to improve their texts. Thus, we can infer that use of the provided support might have facilitated the use of feedback to avoid making erroneous text changes. Additionally, the findings indicate that sense-making support helped students to avoid using incorrect feedback comments for text changes. Overall, these results suggest that sense-making support helped students to reflect on given feedback and, in particular, to think more deeply about which feedback to use for text changes (Schraw 1998). Our treatment check results seem to confirm this picture. Although it was revealed that students did also use erroneous comments, they did so much less frequently compared to correct feedback. Sense-making support might have helped them to become more aware of which comments were correct and which were not. The number of successful changes to errors did not differ significantly between the two conditions. One possible reason why sense-making support was not effective regarding successful changes might be the extra workload with which students in the SMS+ condition needed to deal. It is possible that the time utilized for sense-making support took away time that students in the SMS− condition were able to fully spend for making feedback-based changes in the text. Future studies should ensure that assessees are not overburdened, and that there is sufficient time to make changes.

Second, we were interested in whether sense-making support affects revision skills. We tested whether students receiving sense-making support made improvements in problem detection and problem correction, two crucial subskills of revision. With regard to problem detection, we assumed that the use of sense-making support would help students to reflect on feedback, thereby fostering problem detection. The results partially confirmed our assumptions. There was a trend for students in the SMS+ condition to highlight more text passages. Furthermore, we found that students in the SMS+ condition performed better in terms of labeling errors. Overall, students in both conditions achieved low scores for problem detection. Several reasons might have contributed to this finding: first, the finding might be related to the provision of peer feedback itself. The purpose of peer feedback is to provide the writer with information about problems in the text. As the peer feedback included information that directly pointed out and described errors, students had little incentive to look for errors themselves. According to the assistance dilemma (Koedinger and Aleven 2007), withholding the information (in our case providing no peer feedback and asking the students to look for errors themselves) might have led to more sense making. Future studies should investigate the effect of peer feedback on problem detection in academic writing. Another reason for the low problem detection scores might relate to the type, content, and delivery of feedback. In our study, assessees received local feedback, which pointed directly at the location of the error: feedback comments were incorporated in the text, highlighting the erroneous sentence. Furthermore, feedback included a concise label and description of the error. Thus, the assessee did not need to look for errors because the feedback pointed directly at them. Local feedback makes revision less challenging because restructuring is only necessary at the sentence level, which is less demanding. At the same time, local feedback might lead to less cognitive effort as compared to global feedback, and cognitive effort is important for schema building (Sweller 2005). If a task is less cognitively demanding, reflection might not be as important. Accordingly, it might be argued that our sense-making support was possibly less important. In contrast, global feedback requires the writer to find the error in the paragraph or even complete text. Global feedback points out global text problems, which might be harder to revise and therefore require more in-depth reflection. In other words, an assessee receiving global feedback needs to engage in more problem detection. We suggest that future studies should use global and possibly vague feedback if the goal is to improve problem detection. Another possibility is to provide exemplary feedback, in which errors are

highlighted only for a few instances in the text. In this case, the student still needs to go through the complete draft to find the same kind of errors in other text passages (Funk 2016). Exemplary feedback might therefore leave more scope for students to detect problems themselves.

Our study contributes to research on peer feedback in several ways. We endeavored to tease apart the roles of the assessor and the assessee. Usually, a learner assumes both roles in peer feedback activities: that of the assessor and that of the assessee. While this dual role can be very useful from a pedagogical standpoint, it makes it hard to systematically test effects. Looking at the roles separately and independently from each other might be a first step towards gaining a better understanding of how peer feedback works. Furthermore, we conducted a more robust study by controlling for feedback type and content. In a typical peer feedback activity, feedback is provided by others of equal status, in other words, non-experts. In the present study, feedback was provided by trained tutors, although assessees were told that the feedback came from peers. Although there are good pedagogical reasons for offering "real" peer feedback, its variability causes problems in the experimental setting. For instance, it can range from very concise to very vague, or even conflicting and erroneous. In this study, we controlled for the variability of peer feedback by keeping the type and content of feedback constant. Feedback content was delivered as local feedback only. Furthermore, a feedback comment always included a concise label and description of the error. In order to acknowledge that peer feedback often includes errors, we also used erroneous feedback. In experimental terms, the fact that participants only assumed the role of assessees, and the fact that we controlled for feedback, helped us to reduce variability and the effect of confounding factors and allowed us to infer results more directly from our intervention. Finally, we measured (a) revision skills in a pre-posttest design and (b) feedback uptake by conducting a content analysis. Both outcomes helped us to differentiate between learning and performance, an important difference that is often neglected in peer feedback research.

Several characteristics of our controlled setting limit the generalizability of our findings. As described above, students in our writing activity served as assessees only. Although the students believed that the feedback came from peers, it was actually prepared by trained tutors. While we did include erroneous feedback, our type of feedback nevertheless differed from that provided by actual peers. Furthermore, as we only focused on simple, typical text errors, caution should be exercised when generalizing our findings to more complex text problems. We included erroneous feedback in order to reflect the nature of peer feedback, ensuring the representativeness of our findings. In this study, we focused on feedback-based changes only and did not analyze text changes which assessees carried out independently of received feedback. However, our results show that assessees almost never carried out changes that were not suggested in the feedback. Additionally, we only looked at prose flow, focusing specifically on writing errors, and thus ignored more complex issues such as text coherence or aspects of argumentation.

In sum, we were able to show that peer feedback alone might not be sufficient to make successful changes in a text and improve revision skills. The findings of this study contribute to a better understanding of the conditions under which peer feedback is given and received. Supporting the learner in the moment of feedback provision proved to be effective to some extent and helped to maximize the benefits of peer feedback.

References

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht-Boston- London: Kluwer Academic Publishers.

Allal, L., & Chanquoy, L. (2004). Revision revisited introduction. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision and cognitive instructional processes: Studies in writing* (Vol. Vol. 13, pp. 1–7). Norwell: Kluwer Academic Publishers.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*, 564–577.

Boero, R., & Novarese, M. (2012). Feedback and learning. *Encyclopedia of the Sciences of Learning*, 1282–1285.

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328–338.

Erikson, E. H. (1959). Vertrauen gegen Urmisstrauen. Rezeption und Weiterentwicklung der Psychoanalyse . In F. Baumgart (Ed.), *Entwicklungs- und Lerntheorien: Erläuterungen - Texte - Arbeitsaufgaben 2007*, (pp. 87-93). Bad Heilbrunn: Klinkhardt.

Esselborn-Krumbiegel, H. (2010). Kapitel 3-5. In H. Esselborn-Krumbiegel (Ed.), *Richtig wissenschaftlich schreiben. Wissenschaftssprache in Regeln und Übungen* (pp. 23–74). Paderborn: Schöningh.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessment. *Assessment and Evaluation in Higher Education, 11*(2), 146–166.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57*(4), 481–506.

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication, 37*(1), 16–55.

Funk, A. L. (2016). Unterstützung der Reflexion beim Empfangen von Online Peer-Feedback (Doctoral dissertation). Retrieved from http://www.ub.rub.de/katalog/titel/2980650

Goldin, I. M., Ashley, K., & Schunn, C. D. (2012). Redesigning educational peer review interactions using computer tools: An introduction. *Journal of Writing Research, 4*(2), 111–119.

Graham, S., & Harris, K. R. (2007). Best practices in teaching planning. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 119–140). New York: Guilford Press.

Hammer, R., Ronen, M., & Kohen-Vacs, D. (2010). Stressed yet motivated: Web-based peer assessed competition as an instructional approach in higher education. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th international conference of the learning sciences, Vol. 1* (pp. 65–72). Chicago: IL.

Hayes, J. R. (2004). What triggers revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 9–20). Dordrecht: Kluwer.

Hayes, J. R., & Flower, L. (1986). Writing research and the writer. *American Psychologist, 41*, 1106–1113.

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics. Reading, writing, and language processing* (Vol. Vol. II). Cambridge: Cambridge University Press.

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.

Koedinger, R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344–348.

Kornmeier, M. (2008). Wissenschaftlich schreiben leicht gemacht für Bachelor, Master, Dissertation. Bern, Stuttgart, Wien: Haupt Verlag.

Krämer, W. (1999). *Wie schreibe ich eine Seminar- oder Examensarbeit?* Frankfurth/Main, New York: Campus Verlag.

Kruse, O. (2010). *Lesen und Schreiben*. Wien: Verlag Huter und Roth KG.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536.

MacArthur, C. A., Graham, S., & Fitzgerald, J. (2006). *Handbook of writing research*. New York: Guilford Press.

Moodle Pty. Ltd. (2013). Moodle (Version 2.3.3) [Learning environment software]. Retrieved from https://moodle.org/.

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401.

Newell, A., & Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs.: Prentice-Hall.

Proske, A., Narciss, S., & McNamara, D. (2010). Computer-based scaffolding to facilitate students development of expertise in academic writing. *Journal of Research in Reading, 33*(1), 1–17.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*, 534–574.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (19-30). New York, NY: Cambridge University Press.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. S. R. Segers, F. J. R. C. Dochy, & E. C. Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards* (pp. 55–87). Dordrecht: Kluwer Academic Publishers.

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

Van Gennip, N., Segers, M., & Tillema, H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning & Instruction, 20*(4), 280–290.

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

Zariski, A. (1996). Student peer assessment in tertiary education: Promise, perils and practice. Proceedings of the 5th annual teaching learning forum, (Perth: Murdoch). In J. Abbott & L. Willcoxson (Eds.), *Teaching and learning within and across disciplines* (pp. 189–200). Murdoch: Perth.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*, 329–339.

Dr. Astrid Wichmann. Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, GA 1/53, 44801 Bochum, Email: astrid.wichmann@rub.de

Current themes of research:

Peer Assessment (PA). Collaborative Learning. Inquiry Learning.

Most relevant publications in the field of Psychology of Education:

Wichmann, A., & Timpe, S. (2015). Can dynamic visualizations with variable control enhance the acquisition of intuitive knowledge? *Journal of Science Education and Technology*, 1-12.

Wichmann, A., & Rummel, N. (2013). Improving revision in wiki-based writing: Coordination pays off. *Computers & Education, 62*(0), 262-270.

Wichmann, A., & Leutner, D. (2009). Inquiry Learning: multilevel support with respect to inquiry, explanations and regulation during an inquiry cycle. *Zeitschrift für Pädagogische Psychologie, 23*(2), 117-127.

Dr. Alexandra Funk. Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, GA 1/53, 44801 Bochum, Email: alexandra.funk@rub.de

Current themes of research:

Peer Assessment. Academic Writing. Collaborative Learning.

Most relevant publications in the field of Psychology of Education:

Funk, A. L. (2012, Sept). scaffolding sense-making activities and providing knowledge awareness to enhance students' feedback uptake in online peer assessment. Paper presented at the Seventh European Conference on Technology Enhanced Learning, Saarbrücken, Germany.

Funk, A. L., Wichmann, A., Rummel, N. (2013). Unterstützung der Aufnahme von Feedback im Online-Peer Assessment [Abstract]. Abstractband: Gemeinsam verschieden - 14. Fachgruppentagung Pädagogische Psychologie, 18-19.

Funk, A. L., Wichmann, A., Rummel, N. (2014). Unterstützung der Umsetzung von Feedback und Textüberarbeitungskompetenz im Online-Peer Assessment [Abstract]. Abstractband: 49. Kongress der Deutschen Gesellschaft für Psychologie, 120.

Prof. Dr. Nikol Rummel. Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, GA 1/162, 44801 Bochum, Email: nikol.rummel@rub.de

Current themes of research:

Productive Failure. Collaboration scripts. Collaborative Learning.

Most relevant publications in the field of Psychology of Education:

Rummel, N., Walker, E. & Aleven, V. (2016). Different futures of adaptive collaborative learning support. *Journal of Artificial Intelligence in Education*, *26*(2), 784-795.

Loibl, K., & Rummel, N. (2015). Productive failure as strategy against the double curse of incompetence. *Learning: research and practice*, *1*(2), 113–121.

Rau, M., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *The journal of educational psychology*, *107*(1), 30–46.

Loibl, K., & Rummel, N. (2014). Knowing what you do not know makes failure productive. *Learning and instruction, 34*, 74–85.