Peer feedback on college students' writing: exploring the relation between students' ability match, feedback quality and essay performance

Bart Huisman, Nadira Saab, Jan van Driel & Paul van den Broek

Published online: 15 May 2017.

Submit your article to this journal 🗗

View related articles 🗗

View Crossmark data 🗗

Routledge
Taylor & Francis Group

⊘ OPEN ACCESS Check for updates

Peer feedback on college students' writing: exploring the relation between students' ability match, feedback quality and essay performance

Bart Huisman[a], Nadira Saab[a], Jan van Driel[b] and Paul van den Broek[c]

[a]Graduate School of Teaching (ICLON), Leiden University, Leiden, Netherlands; [b]Melbourne Graduate School of Teaching, The University of Melbourne, Melbourne, VIC, Australia; [c]Department of Educational Science, Leiden University, Leiden, Netherlands

ABSTRACT
There does not appear to be consensus on how to optimally match students during the peer feedback process: with same-ability peers (homogeneously) or different-ability peers (heterogeneously). In fact, there appears to be no empirical evidence that either homogeneous or heterogeneous student matching has any direct effect on writing performance. The current study addressed this issue in the context of an academic writing task. Adopting a quasi-experimental design, 94 undergraduate students were matched in 47 homogeneous or heterogeneous reciprocal dyads, and provided anonymous, formative peer feedback on each other's draft essays. The relations between students' individual ability or dyad composition, feedback quality and writing performance were investigated. Neither individual ability nor dyad composition directly related to writing performance. Also, feedback quality did not depend on students' individual ability or dyad composition, although trends in the data suggest that high-ability reviewers provided more content-related feedback. Finally, peer feedback quality was not related to writing performance, and authors of varying ability levels benefited to a similar extent from peer feedback on different aspects of the text. The results are discussed in relation to their implications for the instructional design of academic writing assignments that incorporate peer feedback.

Introduction

Research on peer feedback in education has expanded in the last two decades. This has increased our knowledge on the reliability and validity of peer feedback in primary, secondary and post-secondary education (Cho, Schunn, & Wilson, 2006; Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Topping, 2009), and with respect to the variables that are important for the design and implementation of peer feedback (e.g., Topping, 1998; van Zundert, Sluijsmans, & van Merriënboer, 2010). However, regarding structural

features such as feedback group composition (see van Gennip, Segers, & Tillema, 2009), there does not yet appear to be consensus on how to optimally match students in terms of ability.

This study focuses on the ability match between students during peer feedback on academic writing in a higher education context. There are three reasons for this focus. First, it seems fair to conclude from the literature that peer feedback can be beneficial to higher education students' learning, and that students can perceive these benefits (Hanrahan & Isaacs, 2001). Students can expect reliable and valid assessments from each other regarding the quality of their work (Cho et al., 2006; Falchikov & Goldfinch, 2000). Also, the process of providing feedback can help students improve their writing performance (Lundstrom & Baker, 2009). Providing peer feedback prompts a reviewer to go beyond mere problem detection, engaging him or her in problem diagnosis and in suggesting solutions (Patchan & Schunn, 2015). Second, being able to provide feedback to peers and utilize feedback from peers can be considered important skills in students' subsequent academic or professional career. Importantly, both students' peer assessment skills and their attitudes towards it can be positively influenced through preparation and practice (van Zundert et al., 2010). Third, academic writing skills are considered important across disciplines and are an integral part of higher education curricula. Given the sometimes large student-to-staff ratios in higher education institutes, however, adequate instructor feedback on academic writing tasks can be a challenge. One aid comes from (web-based) applications that facilitate the peer feedback process (see Luxton-Reilly, 2009, for an overview). With the increasing availability and usability of such applications, the peer feedback process becomes easier to design and implement for academic teaching staff. This may increase the extent to which peer feedback is implemented within academic writing tasks.

Student ability matching

Another benefit of applications that facilitate the implementation of peer feedback is the potential array of possibilities in terms of instructional design. For example, it should be possible to automatically match students on certain criteria, such as ability. Although the potential benefits of student matching have already been discussed in 1998 by Topping, there does not appear to be a clear consensus on whether students should be matched with similar ability peers (homogeneously) or with peers or different ability (heterogeneously).

Regarding the homogeneous matching of students, Topping (2009) prescribes matching students with same-ability peers. In addition, King (1997) argues that beneficial socio-cognitive conflict is more likely between equal peers and that higher level learning is more likely to be accomplished when ideas are exchanged on an equal basis. Also, a mindful, critical appraisal of received feedback may be critical to its effectiveness, which could be stimulated by the uncertainty that the peer's status induces (Gielen et al., 2010). An experimental study by Strijbos, Narciss, and Dünnebier (2010) investigated the relation between peer feedback content and sender's (perceived) competence, on the one hand, and feedback perceptions and revision, on the other hand. Their results suggest that status differences between peers may have negative effects; receiving elaborate, specific feedback from high-ability peers was related to more negative affect and less effective text revision. One possible explanation suggested by the authors is that elaborate, specific feedback from

high-competence peers rendered students to become passive and overly reliant on the feedback they received. These theoretical arguments and empirical findings support the suggestion to match students in a homogeneous manner. However, they do not provide empirical evidence for a direct relation between homogeneous matching with peer feedback and writing performance.

Regarding the heterogeneous matching of students, it has been found that higher ability authors tend to focus more on global issues, detect more problems, and are more likely to use effective strategies for revision than lower ability authors (e.g., Patchan & Schunn, 2015). As a result, they may provide more critical peer feedback than lower ability authors do. Patchan, Hawk, Stevens, and Schunn (2013) differentiated between feedback comments that focused on 'high prose' (flow, logic or insight), 'low prose' (lower level writing issues such as grammar) or 'substance' (issues fixable only with content knowledge). They found that the feedback received by low-ability authors was qualitatively different when they were matched with a high-ability reviewer (heterogeneous match), compared to a low-ability reviewer (homogeneous match). Specifically, low-ability authors received and implemented more 'low prose' and 'substance' feedback from high-ability reviewers. High-ability authors received similar types of feedback, irrespective of reviewer ability. A similar trend was reported for provided solutions. Because feedback containing explicit criticism and suggestions for improvement is likely to contribute to feedback implementation and performance (Nelson & Schunn, 2009), these arguments support matching students heterogeneously. Here too, however, they do not provide empirical evidence for a direct relation between heterogeneous matching with peer feedback and students' writing performance.

Defining student ability and feedback quality

Student ability

Student ability has been defined in different ways in prior research. Generally, a distinction can be made between students' task-related ability (e.g., writing skills) and students' ability to provide peer feedback and/or assess others' work (e.g., use of criteria, see, for example, van Zundert et al., 2010). The current study matched students in terms of task-related ability, that is, their scores on a preceding essay assignment, for four reasons: availability, similarity, proximity and validity. First, this ability indicator was both available and practically applicable. Moreover, comparable ability indices are likely to be available in other higher education institutes. Second, this preceding academic writing assignment was similar to the academic writing assignment central to the current study. Third, the assignment was part of an immediately preceding course in the same curriculum, making it an up-to-date indicator of students' academic writing ability. Fourth and finally, although it is not self-evident that an able writer also is an able reviewer, it is plausible that writing and reviewing ability are interrelated. A rationale for this is provided by Patchan and Schunn (2015), who identified conceptually identical elements between writing and providing feedback on writing: task definition, problem detection and diagnosis, and selection or revision strategy. This overlap in cognitive processes supports the notion that students' ability to write and students' ability to review each other and provide feedback indeed are interrelated, and that high-ability writers can be expected also to be high-ability reviewers.

Peer feedback quality

As is the case with student ability, feedback quality has been defined in multiple ways in the literature. Definitions range from relatively simple categorizations such as holistic feedback versus specific feedback (Lin, Liu, & Yuan, 2001) to more elaborate categorizations such as proposed by Nelson and Schunn (2009) differentiating between summarization, specificity, explanation and scope. The current study adopts the definition of feedback quality as used by van den Berg, Admiraal, and Pilot (2006), which includes the aspect of the text to which the feedback relates (content, structure, style) and the function the feedback comments serve in relation to the text (analysis, evaluation, explanation, revision). An important reason for this choice is that the feedback aspects were aligned with both task instructions and assessment criteria.

In summary, theoretical accounts and empirical findings on how to optimally match students in terms of ability vary and sometimes appear contradictory. To our knowledge, there are no studies that address the direct link between student matching, feedback quality and writing performance. Specifically, there appears to be no empirical evidence that either homogeneous or heterogeneous student matching has any effect on writing performance. Moreover, there is a need for (quasi-)experimental studies investigating the effects of peer assessment (Strijbos & Sluijsmans, 2010). Adopting an exploratory approach, the current quasi-experimental study specifically focuses on the relation between the students' ability match, peer feedback quality and their performance on an academic writing task.

Research questions

This issue was investigated in the context of an essay assignment within a first-year introductory course Education and Child Studies. In this context, student matching was reciprocal, meaning that the students within a particular dyad provided feedback to their dyad member and received feedback from that dyad member. Three main research questions are formulated. Research question 1 is: 'to what extent is student ability in, and dyad composition of reciprocal dyads related to authors' increase in essay performance?' Research questions 2 and 3 explore this relation in more detail. Specifically, research question 2 is: 'to what extent is student ability in reciprocal dyads related to peer feedback quality?' Here, two sub-questions are formulated. The first focuses on reviewers' individual ability: (a) 'what is the relation between reviewer ability and the quality of the peer feedback they provide?' The second sub-question takes into account the interdependence of authors and reviewers within the dyads: (b) 'to what extent does provided peer feedback quality vary between differently composed dyads?' Finally, research question 3 focuses on the relation between peer feedback quality and essay performance: 'to what extent is received feedback quality related to essay performance, and to what extent is this relation moderated by author ability?'

Methods

Participants and procedure

Participants were undergraduate students in a first-year introductory course Education and Child Studies ($N = 220$) at a large research-intensive university in the Netherlands.

In total, 121 students both agreed to participate and submitted all assignments. Ninety-four students were included in the study, as they were part of a dyad in which both students participated. The mean age was 19.8 years ($SD = 1.67$), with 88 females and six males. Students had three weeks to work on a draft essay assignment, followed by one week for peer feedback and another week to produce a final version based on the draft and received peer feedback. Peer feedback was formative, and was provided anonymously and reciprocally within dyads through a virtual learning environment (Turnitin; e.g., Buckley & Cowap, 2013).

Essay assignment, criteria and grading

The essay assignment was instructed to be about 500–750 words excluding references. Students were free to choose one of two essay topics: one in the field of Family Pedagogy ('FP') or one in the field of Educational Sciences ('ES'). For each topic, two scientific articles were provided. The submission of a (serious) draft essay, a final essay and the provision of adequate peer feedback were mandatory course requirements.

Peer feedback guidelines and criteria were provided through a plenary meeting and digital handouts. Essay grades were assigned by the teaching staff according to the following assessment criteria: Content (30%), Structure (20%), Writing style (20%), Referencing (20%), and Presentation and spelling (10%). Within the context of this study, writing style, referencing and spelling were taken together and categorized as elements of Style. Grades ranged from 1 (lowest possible score) to 10 (highest possible score), and grades on the final essay versions were communicated with the students, whereas draft essays were graded for research purposes only.

Essay grading

Draft essays were graded by one trained research assistant, whereas final essays were graded by four teaching assistants. The research assistant was trained by one of the teaching assistants and the first author. Inter-rater agreement between the trained research assistant and the teaching assistant was calculated, based on a subset of 26 draft essays, resulting in high inter-rater agreement ($r = .80$, $p < .001$). Moreover, average scores were similar ($t(25) = 0.07$, $p = .375$). Thus, grades assigned by these two raters provided comparable measures of essay quality. Both graders were blind to the matching condition of the students, but were aware of the manuscripts being drafts or final versions. This was not considered problematic, however, because all graders were instructed to grade the manuscripts using the same standards, and the analyses focused on relative improvement across students (cf. Cho & MacArthur, 2010).

Participant grouping

Dyads were formed by matching students in terms of their ability, defined as students' performance on a similar essay assignment from a directly preceding introductory course in the same curriculum. In the remainder of this study, this ability indicator will be used in relation to students' role both as author and as reviewer.

Within each topic group, students were first rank-ordered on ability, after which they were alternately assigned to one of two conditions (Matching Type): a homogeneous

condition (to be matched with a similar ability peer) or a heterogeneous condition (to be matched with a peer of different ability). Following this procedure, students in the two Matching Type conditions were optimally comparable in terms of ability, both containing high- and low-ability students across the entire range of ability. Next, within the homogeneous condition, dyads were formed by pairing students adjacent on ability. Within the heterogeneous condition, a split-half procedure was conducted to differentiate between higher and lower ability students. A 'moving window' procedure was applied to pair students from the top and bottom half, thereby keeping the ability difference within heterogeneous dyads as constant as possible.

Between topic groups, irrespective of Matching Type, higher ability students in the FP group ($N = 32$, $M = 7.75$, $SD = 0.75$) and the ES group ($N = 15$, $M = 7.64$, $SD = 0.73$) scored similarly on the preceding essay ($t(45) = 0.49$, $p = .629$). For the lower ability students, mean scores for those in the FP group ($N = 30$, $M = 5.54$, $SD = 0.90$) and those in the ES group ($N = 17$, $M = 5.30$, $SD = 1.08$) were similar as well ($t(45) = 0.82$, $p = .419$). Within topic groups, higher and lower ability students significantly differed in both the FP group ($t(60) = 10.47$, $p < .001$) and the ES group ($t(30) = 7.25$, $p < .001$).

Measures and instrumentation

Feedback quality was defined in terms of feedback aspects and feedback functions, in line with van den Berg et al. (2006). *Feedback aspects* concerned the aspects of the text to which the feedback related, distinguishing between content, structure and style. Here, 'Content' referred to clarity of the problem, argumentation and the relevance of the presented information. 'Structure' referred to the internal consistency of the text, such as that between the problem statement, the presented arguments and the discussion. 'Style' referred to grammar, spelling, language use and referencing. *Feedback functions* concerned the function that feedback comments served in relation to the essay in question, distinguishing between 'Analysis', 'Evaluation', 'Revision' and 'Explanation' (Flower, Hayes, Carey, Schriver, & Stratman, 1986; van den Berg et al., 2006). Feedback comments were coded 'Analysis' if they concerned the meaning of the text or the reviewer's perceived understanding thereof. These reviewer comments were regularly phrased as questions, such as 'What do you mean with […]?'. Furthermore, 'Evaluation' referred to feedback comments that included explicit or implicit quality statements. 'Revision' referred to explicit suggestions for improvement of the text, or implicit suggestions for improvement that included at least a direction for a solution (e.g., 'these references are not adhering to APA guidelines'). Finally, 'Explanation' referred to arguments that supported evaluative comments or suggestions for improvement.

Coding procedure

Feedback quality was coded in two steps. First, the peer feedback was coded in terms of feedback aspects. Second, every aspect-segment was also coded as having one or more feedback functions (thus allowing for multiple feedback functions per feedback aspect). Inter-rater agreement for both feedback aspects and functions was determined based on the judgement of two coders. Randomly chosen draft essays on which peer feedback was provided were independently coded for feedback aspects, and agreement was

calculated. Having reached acceptable agreement, the remaining peer feedback was coded for feedback aspects by one coder. This procedure was repeated for feedback functions.

Inter-rater agreement for feedback aspects was calculated based on a random sample of 17 essays. Agreement was moderate for Structure ($k = .59$, 95% CI [.38, .80]) and substantial for Content ($k = .64$, 95% CI [.50, .78]) and Style ($k = .78$, 95% CI [.69, .87]). Using the coded feedback aspects as units of analysis, inter-rater agreement for feedback functions was calculated on another random sample of 10 essays. Agreement was moderate for Explanation ($k = .57$, 95% CI [.33, .81]), substantial for Analysis ($k = .70$, 95% CI [.51, .90]) and Evaluation ($k = .73$, 95% CI [.61, .84]), and almost perfect for Revision ($k = .85$, 95% CI [.76, .93]) (Landis & Koch, 1977).

Analyses

Research question 1: To what extent is student ability in and dyad composition of reciprocal dyads related to authors' increase in essay performance? First, the direct relation between Performance Increase, on the one hand, and Author Ability or Reviewer Ability, on the other hand, was explored. Performance Increase was defined as the difference between an author's score on the draft essay and the final version of the essay. Two linear regressions were performed with Performance Increase as dependent variable and either Author Ability or Reviewer Ability as independent variable. In terms of the ability match between authors and reviewers, an analysis of variance (ANOVA) was performed with Performance Increase as dependent variable and Dyad Composition as independent variable. Dyad Composition was defined as one of four types of ability matches between an author and a reviewer. With homogeneously matched students, this refers either to a match between two higher ability students or to two lower ability students. With heterogeneously matched students, this refers either to a low-ability author matched with a high-ability reviewer or vice versa. In case a significant relation with Dyad Composition was found, post hoc comparisons were performed to identify differences in Performance Increase for the differently composed dyads.

Research question 2: To what extent is student ability in reciprocal dyads related to peer feedback quality? To test the effect of Reviewer Ability on provided feedback quality, a multivariate analysis of variance (MANOVA) was performed with Reviewer Ability (high versus low) as independent variable and Feedback Quality as dependent variable. Feedback Quality was defined as the frequency in which the 12 combinations of feedback aspects (Content, Structure, Style) and functions (Analysis, Evaluation, Explanation, Revision) occurred. See Table 1 for an overview. Subsequent ANOVAs on specific Aspect–Function combinations were performed where appropriate. To test the effect of the ability match between authors and reviewers on feedback quality, a MANOVA was performed with Dyad Composition as independent variable and Feedback Quality as dependent variable. Again, subsequent ANOVAs on specific Aspect–Function combinations were performed where appropriate.

Research question 3: To what extent is received feedback quality related to essay performance, and to what extent is this relation moderated by author ability? To test the effect of feedback quality on essay performance, a hierarchical linear regression analysis was performed with Final Essay Performance as the dependent variable. Author Ability and Draft Essay Performance were included as independent variables in step 1, followed by

Table 1. *Provided* feedback aspects and feedback functions.

		Functions														
		(1) Analysis			(2) Evaluation			(3) Explanation			(4) Revision			(5) Total		
	Reviewer ability[a]	Low	High	Total	Low	High	Total	Low	High	Total	Low	High	Total	Low	High	Total
Aspects	**(A) *Content***	43	64	107	96	135	231	66	104	170	60	118	178	265	421	686
	(B) *Structure*	3	5	8	63	75	138	46	44	90	35	38	73	147	162	309
	(C) *Style*	8	9	17	59	75	134	108	119	227	379	375	754	554	578	1132
	Total	54	78	132	218	285	503	220	267	487	474	531	1005	966	1161	2127

Note: Frequencies based on 1568 feedback Aspects, with multiple feedback functions allowed per aspect.
[a]Low-ability reviewer, $N = 46$, High-ability reviewer, $N = 48$.

received feedback comments on aspects of Content, Structure and Style in step 2. In the third and final step, the interaction terms between, on the one hand, Author Ability and, on the other hand, the received feedback comments on aspects of Content, Structure and Style were added to assess the extent to which the relation between feedback quality and essay performance is moderated by author ability.[1]

Results

Manipulation check

Overall, the preceding essay assignment appeared to be significantly related to the quality of students' draft essays before the peer feedback phase ($r_s = .24$, $p = .020$). With respect to student ability matching, the intention was to create homogeneous and heterogeneous dyads. The average ability difference between students in homogeneous dyads ($N = 25$, $M = 0.12$, $SD = 0.17$) was significantly smaller than that between students in heterogeneous dyads ($N = 22$, $M = 2.27$, $SD = 0.39$), $t(45) = 24.82$, $p < .001$. However, the difference within homogeneous dyads did not equal zero (t (24) = 3.57, $p = .002$). Thus, the two Matching Type conditions differed from each other as intended, although, on average, there still was a minimal difference in ability within homogeneous dyads.

Feedback quantity and quality

For the 94 included draft essays, 1568 peer feedback segments were coded as distinct feedback aspects, averaging 16.68 segments per essay (see Table 1). In terms of the average number of provided feedback segments, higher ability students ($N = 48$, $M = 17.48$, $SD = 12.10$) and lower ability students ($N = 46$, $M = 15.85$, $SD = 8.71$) did not differ ($t(92) = -0.75$, $p = .456$). In general, analytical feedback comments were rare, whereas suggestions for improvement occurred frequently. Students predominantly made such suggestions for improvement about aspects of writing style, however, and to a much lower extent about content-related or structural aspects of the essays. Whereas feedback comments about the content or structure of the text were generally evaluative, feedback comments about stylistic aspects predominantly were suggestions for improvement.

Student ability, dyad composition and performance increase

In general, there was improvement between scores for draft versions ($M = 6.51$, $SD = 1.70$) and scores for final essays ($M = 7.04$, $SD = 0.94$), $t(93) = 2.91$, $p = .005$. Table 2 appears to indicate that the academic writing performance of lower ability students may increase more than that of higher ability students, irrespective of dyad composition. However, performance increase did not depend directly on author ability ($\beta = -0.16$, $p = .117$, $\Delta R^2 = .03$) or reviewer ability ($\beta = -0.02$, $p = .837$, $\Delta R^2 = .00$). Most importantly, dyad composition appeared unrelated to students' essay performance increase ($F(3, 90) = 0.850$, $p = .470$, $\eta_p^2 = .03$). Thus, performance increase was neither related to authors' or reviewers' individual ability, nor related to the composition of the dyad.

Table 2. Essay performance and dyad composition.

Dyad composition	N	Draft essay grade		Final essay grade		Performance increase	
		Mean	SD	Mean	SD	Mean	SD
Low-ability author & Low-ability reviewer	24	6.16	1.46	7.05	0.95	0.89	1.69
Low-ability author & High-ability reviewer	22	6.15	1.95	6.90	1.05	0.75	1.87
High-ability author & Low-ability reviewer	22	7.00	1.45	7.33	1.00	0.32	1.73
High-ability author & High-ability reviewer	26	6.72	1.84	6.91	0.75	0.19	1.79
Average	94	6.51	1.70	7.04	0.94	0.53	1.77

Note: Grades range from 1 (lowest) to 10 (highest).

Student ability in reciprocal dyads and feedback quality

In general, reviewer ability was not directly related to provided feedback quality ($V = 0.10$, $F(12, 81) = 0.77$, $p = .672$, $\eta_p^2 = .10$). However, visual inspection of Table 1 suggests that higher ability reviewers provide more content-related feedback. Specifically, univariate tests suggested that higher ability reviewers provide more content-related suggestions for improvement ($F(1, 92) = 6.23$, $p = .014$, $\eta_p^2 = .06$) and more content-related explanatory feedback ($F(1, 92) = 4.19$, $p = .043$, $\eta_p^2 = .04$). Given the exploratory nature of our research question, however, the risk for Type I errors needed to be addressed. Hence, a Bonferroni correction was applied, after which these results no longer remained significant.

In general, dyad composition also was not related to feedback quality ($V = 0.28$, $F(36, 243) = 0.69$, $p = .908$, $\eta_p^2 = .09$). Only the univariate analysis regarding content-related suggestions for improvement suggested a potential difference between differently composed dyads ($F(3, 90) = 3.44$, $p = .002$, $\eta_p^2 = .10$). On average, 3.00 ($SD = 3.27$) content-related suggestions for improvement were provided within high-ability homogeneous dyads. In contrast, such feedback comments appeared to be less common in low-ability homogeneous dyads ($M = 1.54$, $SD = 1.50$), heterogeneous dyads with high-ability reviewers ($M = 1.82$, $SD = 2.19$), and heterogeneous dyads with low-ability reviewers ($M = 1.05$, $SD = 0.99$). As with the relation between student ability and feedback quality, however, a Bonferroni correction rendered this univariate effect non-significant.

Thus, at first glance, peer feedback quality appears unrelated to either individual reviewer ability or dyad composition. However, a closer look reveals trends suggesting that high-ability reviewers may provide more content-related explanations and suggestions for improvement, and that such suggestions for improvement occur more frequently in homogeneous, high-ability dyads than in dyads of other compositions.

Received feedback quality, author ability and essay performance

Final Essay Performance did not depend on the number of feedback comments that students received on either content-related aspects ($\beta = -0.01$, $t(87) = -0.11$, $p = .911$), structure-related aspects ($\beta = -0.12$, $t(87) = -1.03$, $p = .304$), or style-related aspects ($\beta = 0.00$, $t(87) = 0.00$, $p = .997$) of their draft essay. Moreover, author ability did not significantly interact with feedback comments on these content-related aspects ($\beta = 0.08$, $t(84) = 0.56$, $p = .579$), structure-related aspects ($\beta = -0.15$, $t(84) = -1.08$, $p = .282$), and style-related aspects ($\beta = -0.08$, $t(84) = -0.63$, $p = .529$). See Table 3 for an overview.

Table 3. Hierarchical regression analysis (research question 3).

Step	Variables	B	SE	β	t	sig.
1	(Constant)	6.15	0.53		11.54	0.000
	Author Ability	0.03	0.07	0.04	0.41	0.682
	Draft Essay Performance	0.11	0.06	0.19	1.85	0.067
2	(Constant)	6.35	0.56		11.24	0.000
	Author Ability	0.03	0.07	0.04	0.42	0.677
	Draft Essay Performance	0.10	0.06	0.18	1.74	0.085
	FB_content	0.00	0.03	−0.01	−0.11	0.911
	FB_structure	−0.07	0.07	−0.12	−1.03	0.304
	FB_style	0.00	0.02	0.00	0.00	0.997
3	(Constant)	6.23	0.58		10.73	0.000
	Author Ability	0.04	0.07	0.06	0.49	0.624
	Draft Essay Performance	0.11	0.06	0.20	1.86	0.067
	FB_content	−0.01	0.03	−0.05	−0.37	0.713
	FB_structure	−0.07	0.07	−0.13	−1.01	0.315
	FB_style	0.01	0.02	0.04	0.33	0.746
	Author Ability*FB_content	0.01	0.02	0.08	0.56	0.579
	Author Ability*FB_structure	−0.05	0.05	−0.15	−1.08	0.282
	Author Ability*FB_style	−0.01	0.01	−0.08	−0.63	0.529

Note:
Dependent variable: final essay performance
Step 1: $R^2 = .043$ ($p = .133$).
Step 2: $R^2 = .059$, $\Delta R^2 = .016$ ($p = .687$).
Step 3: $R^2 = .078$, $\Delta R^2 = .019$ ($p = .631$).

Thus, feedback quality did not relate to Final Essay Performance, and no significant moderating (interaction) effect of author ability was found, suggesting that this is the case for all authors irrespective of their individual ability.

Conclusions and discussion

The central aim of this study was to assess whether homogeneous or heterogeneous student matching during the peer feedback phase has an effect on peer feedback quality and students' performance on an academic writing task. In the following section, we discuss our findings in the order of the three main research questions.

Student ability, dyad composition and performance increase

Research question 1 addressed the direct relation between students' ability in reciprocal dyads and authors' essay performance increase. Authors' essay performance increase neither was directly related to their own ability, nor was it directly related to the ability of their reviewing peer. Most importantly, no relation was found between dyad composition and students' essay performance increase. Based on these data, it apparently does not matter how students are matched on ability during the peer feedback phase of an academic writing assignment.

These findings contradict prior research that advocates matching students in any particular way, be it homogeneous or heterogeneous matching. Possibly, the anonymous distribution of essays provided a sufficient degree of uncertainty regarding their peer's status to induce a mindful and critical appraisal of the received peer feedback (Gielen et al., 2010). This may suggest that, conditional on students' (perceived) anonymity, *how* students are matched becomes less relevant, emphasizing the role of student perceptions in the peer feedback process (Strijbos et al., 2010).

Student ability in reciprocal dyads and feedback quality

Research questions 2a and 2b addressed the relation between reviewer ability and provided feedback quality, and the relation between dyad composition and provided feedback quality, respectively. In line with prior research (e.g., Snowball & Mostert, 2013), peer feedback primarily focused on issues relating to writing style. In general, however, reviewer ability was not related to the quality of the provided peer feedback. A closer look suggested that high-ability reviewers may provide more content-related suggestions for improvement and content-related explanations, but this effect disappeared when a Bonferroni correction was applied to control for false positives (Type I errors). Similarly, dyad composition and provided peer feedback quality appeared unrelated, with a possible exception worth mentioning being homogeneous, high-ability dyads: when high-ability students were matched with each other, the number of content-related suggestions for improvement appeared to be higher compared to differently composed dyads. Here too, a Bonferroni correction rendered these differences insignificant. However, because of the rather conservative nature of the Bonferroni correction (it may increase the risk for false negatives, Type II errors), we think these trends deserve a closer look in future research.

If future research would indicate that these trends are reliable, then they may reflect the possibility that high-ability reviewers had a deeper understanding of the assigned theoretical content than the low-ability reviewers. If higher ability reviewers indeed are better at diagnosing problems and selecting strategies for revision (Flower et al., 1986; Patchan et al., 2013), this would explain why they provided somewhat more explanatory feedback and suggestions for improvement on content-related aspects of the texts. These trends could also represent ability differences within a restricted range: in the Dutch higher education context, students typically have completed secondary school at pre-university level, which makes them relatively similar in terms of educational background, age and probably also in terms of writing ability. Regarding the interpretation of these trends, this similar background is important for at least two reasons. On the one hand, it may simultaneously explain why in the current study only non-significant trends were found and justify to consider these trends informative. After all, if these trends become apparent in a (homogeneous) sample that is fairly similar in terms of students' ability, they may become more salient as samples become more heterogeneous (i.e., open or online educational contexts such as MOOCs; see Huisman, Admiraal, Pilli, van de Ven, & Saab, 2016). On the other hand, it suggests that academic teaching staff may not have to worry too much about the ability matching of higher education students in on campus courses, at least when these students are relatively similar in terms of ability.

Received feedback quality, author ability and performance increase

Research question 3 addressed the relation between received feedback quality, essay performance and authors' ability. Authors' essay performance was not related to received peer feedback quality. Specifically, it did not matter whether peer feedback comments focused on content-related, structure-related, or stylistic aspects of authors' drafts. This was the case for all authors, irrespective of their individual ability level.

Whether a student benefits from received (peer) feedback is contingent on his or her mindful reception of, engagement with, and utilization of the feedback (Handley, Price, & Millar, 2011). This study focused directly on authors' summative essay performance, and not on the preceding step of feedback utilization. If we would assume that making revisions based on received peer feedback would generally increase writing performance (although this assumption is debatable, see Flower et al., 1986), our results appear to contradict those of Patchan et al. (2013). Among others, these authors found that, compared to high-ability authors, low-ability authors received and implemented more feedback on 'substance' (issues fixable with content knowledge) from high-ability reviewers. We did not find such a significant relation between dyad composition, content-related feedback and content-related essay performance increase. If anything, a contradicting trend was found in which high-ability authors received more content-related suggestions for improvement than low-ability authors when matched with high-ability reviewers. Possibly, the drafts of high-ability authors were already perceived somewhat better in terms of structure and style, allowing the high-ability reviewers to focus more on content-related aspects.

Limitations and implications

Some remarks are in place. First, we did not take into account students' perceptions regarding the adequacy of the received peer feedback. As such, it remains an open question how the peer feedback was perceived, and how this is related to student ability matching, feedback quality and essay performance. Future research may focus on these relations by incorporating students' responses on questionnaires (e.g., Strijbos et al., 2010) or interviews (e.g., Hanrahan & Isaacs, 2001). Second, as is the case in many studies on peer feedback, the students in this study were simultaneously feedback provider and receiver. Hence, it was not possible to disentangle what the effects on providing versus receiving peer feedback were on students' essay performance. Because the act of providing peer feedback may be as effective as receiving peer feedback (Lundstrom & Baker, 2009), investigating these separate effects in relation to student ability seems an interesting topic for future research.

No differences were found in terms of writing performance for homogeneously and heterogeneously matched students. This suggests that ability matching is not related to students' essay performance and that students may very well be matched randomly. Because random student-matching is a feature of many web-based peer feedback applications, this may simplify at least one decision that academic teaching staff have to make when designing anonymous feedback processes.

Note

1. In our definition of feedback quality, feedback aspects related to either the content, structure or style of the text. This was aligned with the components of the rubric used to assess the final essays (e.g., Content was weighted for 30%, Structure for 20% and Style aspects weighted for 50% in calculating overall essay grades). We are aware that, given these differences between weights, and given that only a single composite final essay grade was available, caution is in place when comparing the impact of these various feedback aspects on writing performance (effect sizes are restrained proportionately to the relative weights attributed to these three

aspects). Therefore, an additional residualizing procedure was conducted in which three separate dependent variables were created: content-related, structure-related and style-related Final Essay Performance. This allowed for a comparison of the separate effects that the feedback comments on aspects of Content, Structure and Style had on students' Final Essay Performance on those particular aspects of the texts. Because no significant relations were found, the different weights of these aspects were not further attended to in the principal analyses.

Acknowledgements

Disclosure statement

References

Buckley, E., & Cowap, L. (2013). An evaluation of the use of Turnitin for electronic submission and marking and as a formative feedback tool from an educator's perspective. *British Journal of Educational Technology*, *44*(4), 562–570. doi:10.1111/bjet.12054

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*(4), 328–338. doi:10.1016/j.learninstruc.2009.08.006

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901. doi:10.1037/0022-0663.98.4.891

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. doi:10.2307/1170785

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision+composition. *College Composition and Communication*, *37*(1), 16–55. doi:10.2307/357381

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, *20*(4), 304–315. doi:10.1016/j.learninstruc.2009.08.007

Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': Investigating the concept of student engagement with feedback. *Oxford Review of Education*, *37*(4), 543–560.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, *20*(1), 53–70. doi:10.1080/07294360123776

Huisman, B., Admiraal, W., Pilli, O., van de Ven, M., & Saab, N. (2016). Peer assessment in MOOCs: The relationship between peer reviewers' ability and authors' essay performance. *British Journal of Educational Technology*. doi:10.1111/bjet.12520

King, A. (1997). ASK to THINK-TEL WHY: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational Psychologist*, *32*(4), 221–235. doi:10.1207/s15326985ep3204_3

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

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*(4), 420–432. doi:10.1046/j.0266-4909.2001.00198.x

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. doi:10.1016/j.jslw.2008.06.002

Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education*, *19*(4), 209–232. doi:10.1080/08993400903384844

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401. doi:10.1007/s11251-008-9053-x

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, *41*(2), 381–405. doi:10.1007/s11251-012-9236-3

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*. doi:10.1007/s11251-015-9353-x

Snowball, J. D., & Mostert, M. (2013). Dancing with the devil: Formative peer assessment and academic performance. *Higher Education Research & Development*, *32*(4), 646–659. doi:10.1080/07294360.2012.705262

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*(4), 291–303. doi:10.1016/j.learninstruc.2009.08.008

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, *20*(4), 265–269. doi:10.1016/j.learninstruc.2009.08.002

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

Topping, K. J. (2009). Peer assessment. *Theory into Practice*, *48*(1), 20–27.

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education*, *11*(2), 135–147. doi:10.1080/13562510500527685

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*(1), 41–54.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270–279. doi:10.1016/j.learninstruc.2009.08.004