Peer Feedback on Writing: The Relation between Students' Ability Match, Feedback Quality, and Essay Performance.

There does not appear to be consensus on how to optimally match students during the peer feedback phase: with same-ability or different-ability peers. The current study explored this issue in the context of an academic writing task. Adopting a quasi-experimental design, 94 undergraduate students provided anonymous peer feedback on each other's draft essays. The relations between students' ability match, feedback quality, and writing performance were investigated. Surprisingly, neither individual ability nor students' ability match directly related to writing performance, and feedback quality did not depend on students' ability match. Also, 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.

Keywords: peer feedback; academic writing; student ability match; feedback quality; higher education.

Theoretical framework and objectives

Research on peer feedback has increased in the last two decades, expanding our knowledge on principles and variables important for the design and implementation of peer feedback (Gielen, Dochy, & Onghena, 2011; Topping, 1998; van den Berg, Admiraal, & Pilot, 2006a; van Zundert, Sluijsmans, & van Merriënboer, 2010). However, regarding the composition of feedback groups, there does not yet appear to be clear 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 the higher education context. There are three reasons for this focus. First, it seems fair to conclude from the literature that, under the right conditions, peer feedback is beneficial to higher education students' learning. Students can, for example,

expect reliable and valid assessments from each other (e.g., Cho, Schunn, & Wilson, 2006; Falchikov & Goldfinch, 2000). Second, being able to provide feedback to peers, as well as being able to utilize the feedback received from peers, can be considered important skills in students' subsequent academic or professional career. Third, academic writing skills are an integral part of higher education curricula. Given the sometimes large student-to-staff ratios in higher education institutes (Ballantyne, Hughes, & Mylonas, 2002), adequate instructor feedback on academic writing tasks can be a challenge. One solution 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.

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 based on a known criterion, such as students' ability. Although the potential benefits of such student matching have already been discussed in 1998 by Topping, there does not appear to be 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, 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.

Regarding the heterogeneous matching of students, 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 (Davies, 2006). Patchan, Hawk, Stevens, and Schunn (2013) found that low ability authors received and implemented more 'low prose' and 'substance' feedback from high ability reviewers, while high ability authors received similar types of feedback, irrespective of reviewer ability. A similar trend was reported for provided solutions. Since feedback containing explicit criticism and suggestions for improvement is likely to contribute to feedback implementation and performance (Nelson & Schunn, 2009), these arguments are supportive of a heterogeneous matching of students.

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 this question by combining measures of peer feedback quality and summative writing performance. The current quasi-experimental study specifically explores this relation between the students' ability match, peer feedback quality, and academic writing performance. There are three main research questions (see Figure 1).

1) To what extent is student ability in, and dyad composition of reciprocal dyads related to essay performance increase after formative peer feedback?

2) To what extent is student ability in reciprocal dyads related to peer feedback quality?

    a. What is the relation between reviewer ability and the quality of the peer feedback they provide?

    b. To what extent does provided peer feedback quality vary between differently composed dyads?

3) To what extent is received feedback quality related to essay performance increase, and to what extent is this relation moderated by author ability?

[Figure 1]

Methods and data

Participants, procedure, and participant grouping

In total, 94 undergraduate students of an introductory course Education & Child studies (N = 220) participated. Students had three weeks to work on a draft essay, one week for peer feedback and one week to produce a final essay. Peer feedback was provided anonymously and reciprocally within dyads through a virtual learning environment (e.g., Rolfe, 2011). Participants were assigned to one of two conditions (Matching Type): a homogeneous condition (with a similar ability peer) or a heterogeneous condition (with a different ability peer).

Essay assignment and grading

The essay was instructed to be about 500-750 words excluding references. The submission of a (serious) draft essay and final essay, as well as the provision of adequate peer feedback, were mandatory parts of the course. Final essays were graded (scale 1-10) based on the following assessment criteria: Content (30%), Structure (20%), and Style

(Referencing, Presentation, and Spelling accumulating to 50%). Drafts and final essays were not graded by the same graders. However, inter-rater agreement was calculated based on a subset of 44 draft essays and was high ($r = .77$, $p < .001$), with average scores being similar ($t(43) = 0.07$, $p = .946$).

Measures and instrumentation

Feedback quality was defined in terms of feedback aspects and feedback functions (see van den Berg, Admiraal, & Pilot, 2006b). *Feedback aspects* concerned the aspects of the text to which the feedback related, distinguishing between 'Content', 'Structure', and 'Style'. *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., 2006b; van den Berg, Admiraal, & Pilot, 2006c). Feedback quality was coded by two coders, with agreement for feedback aspects varying between $k = .59$ ('Structure') and $k = .78$ ('Style'), and agreement for feedback functions varying between $k = .57$ ('Explanation') and $k = .85$ ('Revision').

Analyses

To investigate the direct relation between Performance Increase on the one hand and students' ability on the other hand (*research question 1*), 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. To test the effects of reviewer Ability and Dyad Composition on provided feedback quality (*research question 2*), multivariate analyses of variance (MANOVAs) were performed with reviewer Ability (high/low) or Dyad Composition as independent variable and Feedback Quality as

dependent variable. For *research question 3*, three different hierarchical regression analyses were performed. The dependent variables were residualized (content-, structure-, or style-related) Performance Increase measures. Independent variables were added in two blocks. Main effects were added in block one, being author Ability and feedback quality (the four Aspect-Function combinations for the received feedback aspect in question). The four interactions between author Ability and these particular Aspect-Function combinations were added in block two.

Results

Feedback quality. In general, analytical feedback comments were rare. On the other hand, suggestions for improvement occurred frequently. Students predominantly made such comments 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 (See Table 1).

[Table 1]

Research question 1. Performance increase did not depend significantly on author ability (β = -0.16, p = .117, ΔR^2 = .03), on reviewer ability (β = -0.02, p = .837, ΔR^2 = .00), or on dyad composition ($F(3, 90)$ = 0.850, p = .470, η_p^2 = .03). Thus, performance increase was not related to authors' or reviewers' individual ability, or the composition of the dyad.

[Table 2]

Research question 2. In general, reviewer ability was not directly related to the feedback quality that reviewers provided ($V = 0.10$, $F(12, 81) = 0.77$, $p = .672$, $\eta_p^2 = .10$). Although 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 ($F(1, 92) = 4.19$, $p = .043$, $\eta_p^2 = .04$), a Bonferroni correction was appropriate, rendering these results no longer 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 with respect to content-related suggestions for improvement, a univariate analysis suggested a potential difference between differently composed dyads ($F(3, 90) = 3.44$, $p = .002$, $\eta_p^2 = .10$), with high ability homogeneous dyads producing a higher average of 3.00 ($SD = 3.27$) content-related suggestions for improvement. Again, however, a Bonferroni correction rendered this univariate effect nonsignificant.

Research question 3. Author ability and the received feedback functions (Analysis, Evaluation, Explanation, and Revision) neither predicted authors' content-related essay performance ($F(5, 88) = 1.161$, $p = .335$, $R^2 = .06$), nor their structure-related essay performance ($F(5, 88) = 0.626$, $p = .680$, $R^2 = .03$) or their style-related essay performance ($F(5, 88) = 0.669$, $p = .648$, $R^2 = .04$). Moreover, including the interactions between author ability and received feedback functions did not improve the fit of the regression models (content: $\Delta F(4, 84) = 1.435$, $p = .230$, $\Delta R^2 = .06$; structure: $\Delta F(4, 84) = 0.417$, $p = .796$, $\Delta R^2 = .02$; style: $\Delta F(4, 84) = 2.231$, $p = .073$, $\Delta R^2 = .09$). Thus, no significant moderating (interaction) effect of author ability was found, indicating that authors of varying ability levels did not benefit differently from different feedback functions.

Conclusions scholarly significance

The central aim of this study was to explore the effects of student ability matching on peer feedback quality and students' subsequent academic writing performance. With respect to both peer feedback quality and essay performance, this study suggests that it does not matter how students are matched. These findings contradict prior research that advocates matching students in any particular way, be it homogeneous or heterogeneous matching. Although some trends suggested that high ability reviewers provide more content-related suggestions for improvement and explanations, this effect disappeared after applying a Bonferroni correction. 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. This did not influence students' essay performance, though. Possibly, the anonymous distribution of essays could have provided a sufficient degree of uncertainty regarding their peer's status to induce a mindful and critical appraisal of the received peer feedback (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Yang, Badger, & Yu, 2006). This may suggests that, conditional on students' (perceived) anonymity, *how* students are matched becomes less relevant, emphasizing the role of student perceptions in the peer feedback process (Cheng & Warren, 1997; Strijbos et al., 2010).

If future research confirms that ability matching is unrelated to students' essay performance, students may very well be matched randomly without taking into account their writing ability. Since 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 peer feedback processes.

References

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education, 27*(5), 427-441. doi: 10.1080/0260293022000009302

Cheng, W. N., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*(2), 233-239.

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: Doi 10.1037/0022-0663.98.4.891

Davies, P. (2006). Peer assessment: Judging the quality of students' work by comments rather than marks. *Innovations in Education and Teaching International, 43*(1), 69-82. doi: 10.1080/14703290500467566

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: 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., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137-155. doi: 10.1080/02602930903221444

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: http://dx.doi.org/10.1016/j.learninstruc.2009.08.007

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

Rolfe, V. (2011). Can Turnitin be used to provide instant formative feedback? *British Journal of Educational Technology, 42*(4), 701-710. doi: 10.1111/j.1467-8535.2010.01091.x

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: http://dx.doi.org/10.1016/j.learninstruc.2009.08.008

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. (2006a). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341-356. doi: 10.1080/03075070600680836

van den Berg, I., Admiraal, W., & Pilot, A. (2006b). 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 den Berg, I., Admiraal, W., & Pilot, A. (2006c). Peer assessment in university teaching: evaluating seven course designs. *Assessment & Evaluation in Higher Education, 31*(1), 19-36. doi: 10.1080/02602930500262346

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: DOI 10.1016/j.learninstruc.2009.08.004

Yang, M., Badger, R., & Yu, Z. (2006). A Comparative Study of Peer and Teacher Feedback in a Chinese EFL Writing Class. *Journal of Second Language Writing, 15*(3), 179-200.

Table 1. *Provided* feedback aspects and feedback functions

		Functions															
		(1) Analysis			*(2) Evaluation*			*(3) Explanation*			*(4) Revision*			*(5) Total*		
	Reviewer ability[1]	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

[1] = Low ability reviewer $N = 46$, High ability reviewer $N = 48$ — Frequencies based on 1580 feedback Aspects, with multiple feedback functions allowed per aspect.

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

Grades range from 1 (lowest) to 10 (highest)
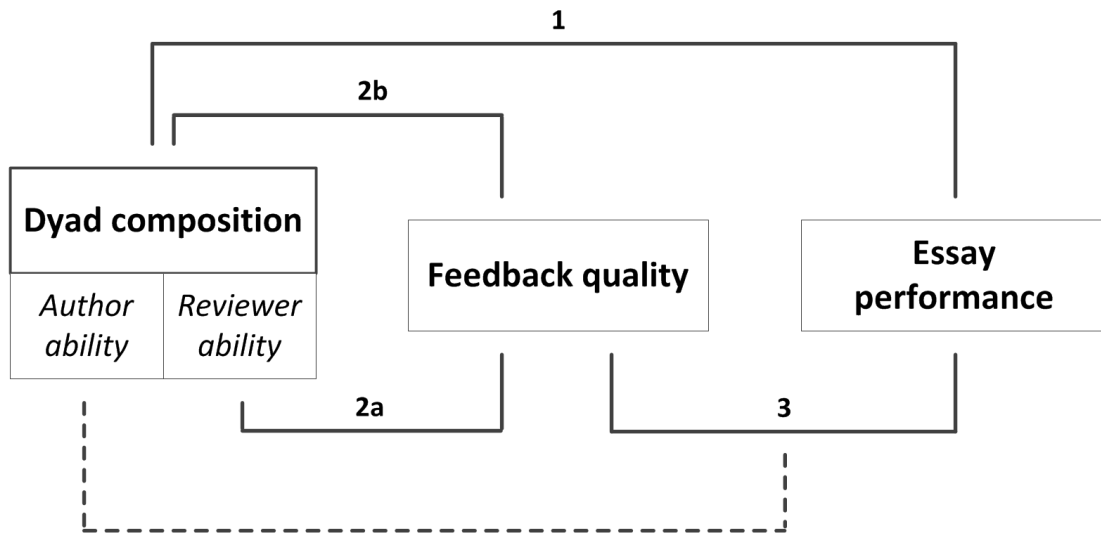
Figure 1. Graphical representation* of research questions

* = Straight lines refer to main research questions, dashed line indicates moderation.