Dear Pat: The role of error detection problems for learning effective peer feedback

Elizabeth S. Charles, Dawson College, echarles@dawsoncollege.qc.ca
Michael Dugdale, John Abbott College, michael.dugdale@johnabbott.qc.ca
Kevin Lenton, Vanier College, lentonk@vaniercollege.qc.ca
Chao Zhang, McGill University, chao.zhang2@mail.mcgill.ca
Rhys Adams, Vanier College, adamsr@vaniercollege.qc.ca

Abstract: While the positive potential of peer feedback has been widely studied, it is difficult to isolate the effect that the providing of feedback to a peer has on a learner due to the variability in source material (the peer's submitted work). This variability can lead to unequal learning opportunities for the students. In this study, we explore the use of Error Detection Problems (EDPs). In these activities, students are given the same source material: a worked problem produced by a fictitious peer 'Pat'. We show that engaging in these activities can help students gain the positive impacts expected of authentic peer feedback exercises, while minimizing the effects of source material variability. Furthermore, we argue that such activities may provide an opportunity for researchers to better understand how students learn to give effective feedback and for CSCL to design new forms of learning scripts.

Introduction

Research shows that peer feedback positively enhances students' learning (e.g., Falchikov, 2003; Nelson & Schunn, 2008) and is effective for improving students' domain-specific skills across a variety of disciplines (van Zundert et al., 2010). It is a reciprocal process involving two components: giving and receiving feedback. To date, the literature has not always unpacked the differences between the two processes (Li & Grion, 2019) — i.e., whether or not they trigger different mechanisms that account for the learning improvement reported. Research on the role and impact of peer feedback typically has focused on the topic of improving writing skills (e.g., Huisman et al., 2019) leaving open the question of how peer feedback might support learning of other skills such as problem solving. Additionally, there is a paucity of carefully controlled studies, this includes the variability in how studies have identified conditions, methods, and outcomes of peer feedback (van Zundert et al., 2010).

While much has been written about the positive effects that receiving peer feedback has on student learning, this study focuses on the impact that providing feedback has on student learning and performance. Giving feedback to a peer is, we hypothesize, an important mechanism through which students can develop self-regulatory skills by learning how to monitor and evaluate their growing knowledge (Ferguson, 2011). In doing so, they gain the skills necessary to better evaluate their own work (Nicol et al., 2014). This self-regulatory and agentic action improves with practice (Brown et al., 2016) and training (Alqassab et al., 2018).

Investigating how students learn to give effective feedback is, however, complicated in authentic peer feedback contexts due to the wide variability of source material (their peer's work they are to assess). Our approach targets a condition that has not been the focus of previous studies — i.e., removing this variability in the source condition by providing all students with the same material ostensibly produced by a fictitious peer in what we call error detection problem (EDP) questions. We argue that this approach helps to better reveal what aspects of the error students are attending to when responding and giving feedback to peers. This strategy holds potential for designing computer-supported collaborative learning environments (CSCL) that leverage the use of peer feedback exercises to promote a deeper learning of the course content.

This paper describes a design-based-research (DBR; Anderson & Shattuck, 2012) aimed at investigating the affordances of learning repertoires built on three principles from the literature: scripts (EDP questions focused on giving peer feedback); scaffolding student engagement patterns (asynchronous peer instruction script); and, in some instances, training on giving peer feedback. We report on the quasi-experimental study conducted in the final iteration of the 3-semester DBR, as well as provide insights from the process data illustrating how students responded to the EDP component of the repertoires.

Background

Engaging students in collaborative tasks such as peer assessment that include providing feedback to each other is a central part of pedagogies grounded in social constructivist theories of learning — popularly referred to as student-centered active learning (AL). Peer assessment can be viewed as involving the joint construction of knowledge through exchange of dialog (Falchikov & Goldfinch, 2000), and learning from this mutual reflection and judgement on performance (Strijbos & Sluijsmans, 2010). Research shows that peer assessment can promote the development of higher order cognitive and metacognitive skills such as critical thinking, problem solving, and decision-making (King, 2002). It can be a complete activity or a component of an elaborated pedagogical pattern. Complicating matters, the literature uses different terms to describe this process — peer assessment, peer review, peer evaluation, peer feedback. This lack of clarity makes it difficult to compare studies. Here, we consider peer assessment as an overarching process which includes peer feedback.

Peer feedback can have a positive effect on learning (Topping, 1998). Research shows that the type of feedback students gave each other might be different compared to the explanations they receive from a teacher (Yang et al., 2006). The cognitive and metacognitive skills required to use feedback, as an effective tool for knowledge construction, is not automatic and need to be learned (King, 2002). Training of peer assessors is linked to the success and effectiveness of peer feedback (e.g., Sluijsmans et al., 2002) including improvement of the psychometric qualities (reliability and validity) of the feedback. In addition, it appears to be related to improvements in students' thinking as well as attitudes towards the process of peer assessment itself (van Zundert et al., 2010). Some of these training programs have been incorporated into collaborative scripts to support the learning of peer feedback skills (Gielen et al., 2010).

Scripts have long been used to support and scaffold the cognitive load associated with high-level cognitive and metacognitive demands (King, 2002). Collaborative scripts are intended to guide the learner in social activity including defining roles, establishing ways for turn-taking, work phases, and so on (Dillenbourg & Jermann, 2007). In addition, scripts can represent models at different levels of granularity: (1) macro, pedagogical models, and (2) micro, dialogue models embedded in the environment (Dillenbourg & Hong, 2008). How these levels can be brought together under one intervention is a question to be examined.

In recent years web-based tools that support peer feedback have appeared. For instance, Perusall is an online platform in which students annotate, publicly, a shared reading. The quality of their annotations is assessed by other students in the class (Miller et al., 2016) as part of a pedagogical AL model. Another popular peer assessment platform is Peerceptiv (previously SWoRD; Cho & Schunn, 2007), in which students upload their completed assignments and anonymously assess the work of other students, according to a teacher-supplied rubric, as part of a micro peer feedback dialogue model. Supporting the cognitive and metacognitive skills involved in peer feedback may require more explicit scaffolds and scripts (Vogel et al., 2017).

Design of the learning repertoires

Our training and scaffolding model design brings together three repertoires to engage the learners in the rehearsal and repetition of the peer feedback skills: (1) a problem-solving dialogue model for peer feedback (the EDP script); (2) a pedagogical method promoting metacognitive engagement (the peer instruction script embedded in myDALITE), and (3) a one-time training module on how to give feedback. Because of page limits we provide only a brief description here of the myDALITE asynchronous peer instruction script was designed to scaffold and privilege students' self-explanations and reflective comparisons (Charles et al., 2014). Briefly, the script involves 3 tasks: (1) answer the multiple-choice question and provide a written explanation — i.e., give feedback, (2) compare the feedback given to those provided by peers (taken from the platform's database), and (3) take the opportunity to choose one of these feedback explanations, if it is perceived to be better or more complete. We use myDALITE to embed the EDP questions, as part of the designed peer feedback repertoires.

Problem solving dialogue model for peer feedback

Problem solving is a major domain-specific skill in STEM education and particularly in learning physics. Error detection as instruction has been used in math education (Heemsoth & Heinze, 2016) but less so in physics. A continuing issue is that students make the common conceptual mistakes when solving problems, they confuse the conceptual understanding (the why) and the procedural understanding (the how). In this context, EDP are questions framed to look like typical solutions produced by novice learners, that contain alternative conceptions (e.g., Sequeira & Leite, 1991). Students, in answering this question type, are expected not only to identify the mistake, but to also explain to others (peers) how to correct the error. In doing so, they are invited to reflect on both the procedural and conditional knowledge.

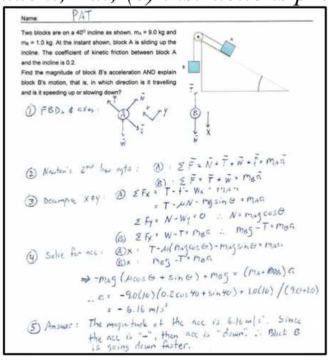
The literature reports that this form of error identification can be more effective under certain conditions, compared with other problem-solving tasks. For example, Große and Renkl (2007) found that students with medium-high initial prior knowledge learned best from a mix of correct and error-detection problems, while students with low prior knowledge benefitted more from studying correct worked examples alone rather than from a mix of correct worked problems and problems with errors, even if the specific errors were highlighted. Importantly, the inclusion of explicit feedback steps seems to play a role. Stark, Kopp, and Fischer (2011) showed a benefit for all students from a mix of correct and error-detection problems when giving explicit feedback and comparing/contrasting. The effectiveness of error identification has also been established compared with computer-based problems that provide immediate feedback on individual steps. In these studies, the treatment groups performed better on a delayed posttest than the control group (McLaren et al., 2016). While these studies provide important insights into the process and use of error detection, none of them explicitly combined this with the provision of peer feedback — this current study does so.

Content of the error detection problem (EDP)

An example of an EDP and peer feedback question is shown in Figure 1. The problem is broken up into sections that allow students to narrow down where the error is exactly (the error in this case concerns the description of the motion in part numbered 5). Students are presented with the solution to a physics problem written by Pat. Pat's solution contains zero, one, or multiple errors which can be algebraic, procedural, or conceptual, depending on the learning objectives. Students are asked to identify and explain how to correct the error(s), and to explain as though they were providing feedback to this classmate. This requires a more careful reading and understanding of the task and helps students to avoid making similar errors and misconceptions in the future. In this designed intervention, we introduce students to a fictional classmate 'Pat', the author of frequently flawed solutions — i.e., protagonist in the EDP.

Figure 1

Example of an EDP and peer feedback question: (a) the error containing solution written by a fictional student, Pat; (b) instructions provided to students.



(a)

"Your best friend Pat has completed a quiz on dynamics and has a copy for you to review. The solution has one or more errors. If there is only one error, help Pat by stating where the error is found (parts 1–5). If there are multiple errors, state which is the most important error and justify your choice. Explain to Pat how the solution can be improved, or the error(s) resolved. Provide your response as if you are speaking to Pat; you want to be a good peer so provide constructive feedback."

(b)

N.B. The error illustrated in this example involves the sign of the acceleration. The role of the sign is often misunderstood by students completing coupled-motion problems such as this.

Training module on how to give feedback

Based on the peer feedback literature, as part of the intervention we designed a 30-minute training module. It included clear statements about three features of giving feedback: (a) identify where the problem/mistake/issue is located, (b) describe what may have gone wrong and why, and (c) explain what actions might be taken to correct/improve the future work. It also specifically addressed the importance of using a positive tone that is respectful and focuses on the strategies and not just getting things right. This training included three practical exercises involving selection, ranking, and explanation. The module was given in class as a 30-minute exercise, in week four of the semester and after the students had had three homework assignments (four questions on myDALITE) with opportunities to provide their fictional peer, 'Pat', with feedback.

Methods

Research design, setting, and participants

This DBR was part of a 3-year study that included three cycles, all involving the same physics teacher but different cohorts of students at a junior college in Québec, Canada. The first and third (final) iterations, Fall 2018 and Fall 2019 respectively, were carried out in an introductory physics course (Mechanics) typically taken by first-year students enrolled in a 2-year pre-university Science Program (equivalent to grade 12 and first-year university). The second iteration, Winter 2019, was conducted in the second required physics course (Waves and Modern Physics) of that Program.

The first iteration, Fall 2018, provided an opportunity to test out the designs of the EDP questions within the myDALITE platform, with four questions produced and tested. We also explored the scripting and placement of the peer feedback training within the curriculum sequence. It included a small quasi-experimental component which showed some differences between the treatment group and control groups but also identified refinements in the EDP question design (Adams et al., 2019). We identified quantity (word count) and quality (content score) of student responses as productive measures. In addition, we used Latent Semantic Analysis (LSA; Dumais, 2004) to assess the within-group similarity or normativity of student feedback. The second iteration, Winter 2019, allowed us to further explore the design requirements for the EDP format and how to prepare students to take up the affordances for learning of the question type.

The final iteration, Fall 2019, is the implementation on which we report in this paper and includes both the results of a quasi-experimental study as well as process data collected from a full semester 16-week implementation of the EDP questions (see Table 1). Participants include one treatment group (T1, N=36), one partial treatment group (T2, N=31) and a control group (C1, N=28). Their equivalence was established by the pretest Force Concept Inventory (FCI) scores that show no statistical differences in their knowledge of the concepts of Newtonian force and its relationship to motion. This allows us to meaningfully compare their performance on a common EDP final exam question. We elaborate on the data collected in Table 2.

Table 1
Interventions by Group

F 2019	wk.3	wk. 4	wk. 6	wk. 10	wk. 11	wk. 12	wk. 13	wk. 14	wk. 15
T1	EDP-1, 2	EDP-3	EDP-4	EDP-5	EDP-6	EDP-7	EDP-8	EDP-9	EDP-10
T2					EDP-6	EDP-7	EDP-8		

Table 2
Assessments, Measures, and Questions of the Quasi-Experiment

Activity/Assessment	Measure	Question
Force Concept Inventory	pre/post concept knowledge	Are the groups comparable?
EDP_final exam_ Q8d	Word counts	Do students in the different groups write quantitatively differently?
EDP_final exam_ Q8d	correctness of content	Do students in the different groups write qualitatively differently about the content?
EDP_final exam_ Q8d	LSA similarity measures	Have students adopted a shared understanding of what constitutes effective feedback?

Results

Quasi-experimental study Fall 2019

A common EDP-based question (Q8d) was posed to the students in each group on the Fall 2019 final examination of a 1st-semester college physics course (Mechanics). The responses were compared between Treatment T1, partial Treatment T2 and Control C1 groups to confirm that the training and practice of giving feedback to a fictitious peer over the course of a semester had a measurable impact on performance of this task at the end of the semester.

While the groups performed similarly on the traditional exam questions, we focused our attention on this EDP-like question. The question involved writing a response to 'Pat' and we examined the word count and the quality of the physics content (assessed using a coding rubric) for these responses, as well as whether the treatment group had developed a more normative idea of how to provide effective feedback to a peer (assessed using LSA). Note that because the unit of analysis is the rationale, the N value for the results differs from that of the class size as some students in each group did not provide a response.

Table 3
Word Count, Physics Score, and LSA Similarity of Feedback to a Fictional Peer on the Final Exam

Group	N	Word Count	Physics score	LSA Similarity
T1	25	M = 96.56, SD = 39.12	M = 1.86, SD = 0.55	M = 0.903, SD = 0.04
T2	22	M = 69.45, SD = 37.50	M = 1.00, SD = 0.93	M = 0.866, SD = 0.07
C1	21	M = 55.05, SD = 29.36	M = 0.86, SD = 0.65	M = 0.827, SD = 0.07

The responses to the common question revealed the effects of the treatment on a Pat-cued question on the final exam (Table 3). Between-subjects ANOVA was conducted to determine what, if any, effect the treatments had on the feedback length (number of words), quality (physics score), and normative style (LSA similarity) of student responses.

Feedback length

In the case of rationale length, the effect of the treatment group was significant: $F(2,65) = 8.04$, $p = 7.6 \times 10^{-4} < 0.05$. Tukey's post-hoc test revealed that treatment group T1 wrote significantly more words than group T2 ($p_{\text{adj}} = 0.32 < 0.05$) and control group C1 ($p_{\text{adj}} = 6.4 \times 10^{-4} < 0.05$). Although students in the T2 condition did write more than those in C1, the effect was not significant at the $\alpha = 0.05$ level.

Quality of physics analysis

Although the above differences in word count might be attributable to the overhead of directly addressing 'Pat', we also found that the quality of physics responses was improved in the treatment group. Essential elements of the response were coded with a positive value, while incorrect elements (e.g., incorrectly associating the normal force with gravity) were coded with a negative value. The codes were summed for an overall score reflecting the quality of the response's physics correctness and completeness.

Between-subjects ANOVA found that the effect of treatment group was significant: $F(2,65) = 13.34$, $p = 1.4 \times 10^{-5} < 0.05$. Tukey's post-hoc test found the physics content of group T1 to be significantly higher than those of either T2 ($p_{\text{adj}} = 3.7 \times 10^{-4} < 0.05$) or the control C1 ($p_{\text{adj}} = 4.2 \times 10^{-5} < 0.05$) although, again, the differences between treatment group T2 and C1 were not statistically significant. This positive effect of treatment may arise from the act of directly responding to the peer, rather than responding to the teacher. We hypothesize that this entails a deeper engagement with the task, or developing a sense of the 'other', which we discuss later.

Similarity of responses

Comparing the in-group homogeneity of responses to the EDP exam question was done by way of LSA similarity measures. These scores are based on the 400-dimension Baroni word-vector embedding space (Baroni, 2014). Within each treatment group, each response was encoded as a sum of word vectors in this semantic space, and the cosine similarity was computed for each of the $(N^2 - N)/2$ unique pairs of responses.

The distributions of these similarity measures (Figure 2) were then compared between groups. Once again, between-subject ANOVA indicated a significant dependence of LSA similarity measures on treatment group: $F(2, 738) = 97.93, p < 2 \times 10^{-16}$. Moreover, post-hoc Tukey analysis found significant ($p \approx 0$) differences between all pairs of treatment groups.

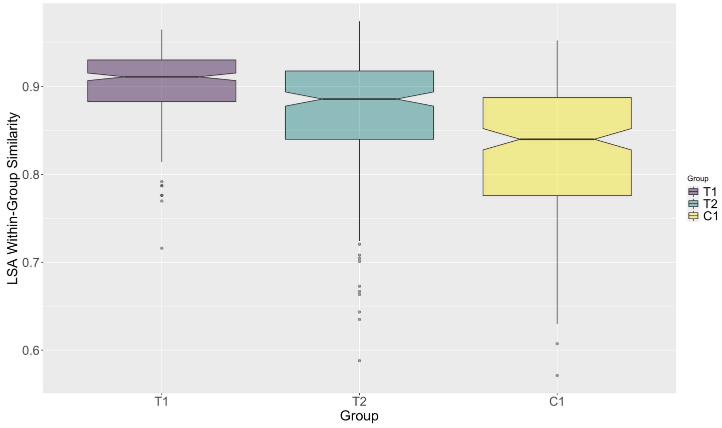
The higher in-group LSA similarity of the treatment group reflects a greater degree of homogeneity in student responses to the exam question. The treatment group T2 also exhibited a significantly elevated degree of homogeneity, although less pronounced than the full treatment. This is notable, as it suggests that even a small EDP-based intervention can foster more normative student writing.

Summary of quasi-experimental results

In summary, students in the full treatment group T1: (a) wrote significantly longer responses for their answers (when prompted), (b) wrote significantly more complete (i.e., richer) responses, and (c) exhibited an increased level of homogeneity in their responses, indicating the presence of a stronger class norm as to what constitutes a good rationale than did students in the control group. Moreover, we observe that even a small EDP-only intervention (T2) results in more normative student responses.

These results are unsurprising, given the previously reported results on the effectiveness of giving peer feedback. However, we argue that they support the claim that simulating feedback to an ostensible peer offers many of the advantages of providing authentic peer feedback, while mitigating the high variability of the starting condition (the peer's submitted work) that can lead to inequitable learning conditions and complications in measuring the impact of feedback interventions. Moreover, it affords us the ability to examine how students learn to give feedback over the course of a semester.

Figure 2
LSA Similarity Scores for Responses to Final Exam Question 9b.



Experience of the EDP 16-week intervention

Over the course of the semester, the treatment group (T1) took part in over 10 EDP-based activities on the myDALITE platform. The student responses offer an interesting window into how students learn to give effective feedback, because comparison between groups and between individual students is greatly simplified by the fact that all the students are responding to the same starting conditions.

Constructive and supportive feedback

Of note in the student responses to EDPs over the course of the intervention is the clear effort being made by the students to understand the "other". While the students were urged to be supportive and constructive in their feedback, many eventually took this further, trying to understand how and why Pat's misunderstanding could have arisen, at the same time being surprisingly sensitive to Pat's "feelings". Consider the following exemplars of responses to the exam question (exemplars are the responses that loaded most heavily on the 1st principal component of their respective group's LSA similarity matrix).

> Hey Pat, you did a good job with all the forces acting on the body in the y-axis however you made a mistake on the forces on the x-axis. First, centripetal force is not actually a force, I know its weird, but its like acceleration and velocity, where you don't include them on the FBD but they do affect the body. Next is your normal You have it in the write [right] axis but in the wrong direction. Remember that it is perpendicular to the surface which the force is being applied. Nice Try but keep practicing. (T1 exemplar)

> The centripetal force should not be in your FBD. It is a resultant force in a circular motion. It is caused by the addition of the forces. Also, the normal force is in the wrong direction. If the person stays on the wall, he/she must be pushed to the center of the circle (centripetal acceleration). Therefore, the only force that is acting on the same axis as the centripetal acceleration is the normal force. So to have a resultant acc that is towards the center, the normal force should also be towards the center. (T2 exemplar)

> First, he can start by drawing an actual FBD. He would simply need to draw a rough out line of the rider and the wall he is on. It is also preferable to give a title to the diagram. Secondly, the centripetal force should not be included, since it is actually the resultant of the addition of the other forces. The normal force should be in the other direction since in this scenario, it is the normal force that causes the centripetal acceleration. Lastly, he should always add the direction of the motion and the acceleration, if there is any. In this case, it should be in the same direction as the centripetal force. In the end, the FBD should look like: [redrawn FBD with normal force correct]. (C1 exemplar)

In analyzing these exemplars, we draw attention to the differences in whom the feedback is directed. This T1 student directed directly to 'Pat', T2 student refers to 'Pat' indirectly through the imperative tense (your), while the C1 student directs the feedback as if evaluating, in the third person (suggesting the response was directed to the teacher). We note that the majority from the T1 group included many more instances (e.g., "you did a good job", "Nice try, but keep practicing") that we interpret as the expression of empathy for their fictitious classmate. Furthermore, we observed these features developing in the T1 group over the course of the 16-week intervention. We would like to explore these further as they raise interesting questions of whether developing empathy can be a way of increasing students' engagement in peer assessment activities.

Discussion and Conclusions

Results of the quasi-experimental study conducted on the last iteration of a 3-year DBR study show that the three repertoires intervention (EDP, asynchronous peer instruction and training) resulted in achieving some positive outcomes. Students in the treatment groups (T1 & T2) performed better on our assessment measures compared to a control group (C1), which supports previous work on the effectiveness of giving feedback to peers (e.g., Li & Grion, 2019). Additionally, participants were positive about the peer feedback approach, also seen in other studies (e.g., van Zundert et al., 2010). However, the three repertoires were not explicitly examined on their own in our full intervention. We therefore do not know if the EDP script alone is responsible for the increased learning, or whether coordination with the asynchronous peer instruction script was essential.

What we do know is that the T2 (minimal, EDP-only) treatment showed an increase in response homogeneity suggesting that training may not be essential or account for the clear difference between the full-treatment T1 and control C1 groups. Although not statistically significant with our small sample sizes, the T2 treatment's increased word count and physics scores suggest that even EDP-only treatments may improve student learning and performance. Further study is needed.

Our most notable finding relates to the participants' response to the fictional peer 'Pat'. Further to Yang and colleagues (Yang et al., 2006) — students' response to feedback from peers is different than feedback from teachers — our study reveals more about such differences. Not only did our treatment group T1 provide more meaningful feedback to the fictional peer, they also wrote more and wrote better physics on final exam questions that prompted for feedback. This finding, addressing peers rather than the teacher, should be examined further for its potential to increase student engagement in CSCL scripts. In addition, the data consists of phrasing associated with empathy, mentioned above. With a majority of students displaying such affect this finding suggests that a community of learners' identity was being formed, arguably a desirable outcome for collaborative learning.

Lastly, our results suggest simulating feedback to an ostensible peer offers many of the advantages of providing authentic peer feedback, while mitigating the high variability of the starting condition (the peer's submitted work) that can lead to inequitable learning conditions and complications in measuring the impact of feedback interventions. This consideration could guide the design of future CSCL interventions as they take up the challenge of greater equity and inclusion in our scripts and scaffolds. And, it affords us the ability to examine how students learn to give feedback over a semester, opening new ways to explore this form of instruction.

References

Adams, R., Jackson, P., Lenton, K., Dugdale, M., Whittaker, C., Lasry, N., & Charles, E. S. (2019, July). Error detection tasks and peer feedback for engaging physics students. *In Fifteenth Conference on Education and Training in Optics and Photonics: ETOP 2019 (Vol. 11143, p. 111433J)*. International Society for Optics and Photonics. http://bitly.ws/q7e6

Alqassab, M., Strijbos, J. W. & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education*, *33*, 11–30. DOI: 10.1007/sl0212-017-0342-0

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research?. *Educational researcher*, *41*(1), 16–25. https://doi.org/c3kc

Baroni, M., Dinu, G. & Kruszewski. G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 238–247. https://aclanthology.org/P14-1023.pdf

Brown, G. T., Peterson, E. R. & Yao, E. S. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology*, *86*, 606−629. https://doi.org/gf9dsc

Charles, E.S, Whittaker, C., Lasry, N., Dugdale, M., Lenton, K., Bhatnagar, S., Guillemette, J. (2014). Taking DALITE To the Next Level: What have we learned from a web-based peer instruction application. In Polman, et. al., (Eds.). *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS) 2014*, Volume 2, pp. 982–987. https://repository.isls.org//handle/1/1225

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, *48*(3), 409–426. https://doi.org/bzr6gz

Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, *3*(1), 5–23. https://hal.archives-ouvertes.fr/hal-00692019

Dillenbourg, P., & Jermann, P. (2007). *Designing integrative scripts. In Scripting computer-supported collaborative learning* (pp. 275–301). Springer, Boston, MA. https://doi.org/dg52cb

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, *38*(1), 188–230. https://www.learntechlib.org/p/97537/

Falchikov, N. (2003). *Learning together: Peer tutoring in higher education*. Routledge.

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. https://doi.org/frqtkh

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, *36*, 51–62. https://doi.org/ffnck5

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. https://doi.org/fbw9fw

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes?. *Learning and instruction*, *17*(6), 612–634. https://doi.org/dc75ps

Heemsoth, T., & Heinze, A. (2016). Secondary school students learning from reflections on the rationale behind self-made errors: A field experiment. *The Journal of Experimental Education*, *84*(1), 98–118. https://doi.org/hpqg

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment & Evaluation in Higher Education*, *44*(6), 863–880. https://doi.org/gnmzqv

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, *41*, 33–39. https://doi.org/bnkbdp

Li, L., & Grion, V. (2019). The Power of Giving Feedback and Receiving Feedback in Peer Assessment. *All Ireland Journal of Higher Education*, *11*(2).

McLaren, B. M., van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worke examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior*, *55*, 87–99. https://doi.org/f76dt9

Miller, K., Zyto, S., Karger, D., Yoo, J., & Mazur, E. (2016). Analysis of student engagement in an online annotation system in the context of a flipped introductory physics class. *Physical Review Physics Education Research*, *12*(2), 020143. https://doi.org/gmkv54

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional science*, *37*(4), 375–401. https://doi.org/b83vbt

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment and Evaluation in Higher Education*, *39*, 102–122. https://doi.org/gcphzq

Sequeira, M., & Leite, L. (1991). Alternate Conceptions and History of Science in Physics Teacher Education. *Science Education*, *75*(1), 45–56. https://doi.org/d3wssp

Sluijsmans, D. M., Brand-Gruwel, S., & van Merriënboer, J. J. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*(5), 443–454. https://doi.org/cnfr4v

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and instruction*, *21*(1),22–33. https://doi.org/c63vrg

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and instruction*, *20*(4), 265-269. https://doi.org/c7dghz

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, *68*(3), 249–276. https://doi.org/d8cq8b

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.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477–511. https://doi.org/gf4tdq

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. https://doi.org/cwjxvj

