Using and rejecting peer feedback in the science classroom: a study of students' negotiations on how to use peer feedback when designing experiments

Jens Anker-Hansen & Maria Andrée

Published online: 04 Jan 2019.

Submit your article to this journal

Article views: 651

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Using and rejecting peer feedback in the science classroom: a study of students' negotiations on how to use peer feedback when designing experiments

Jens Anker-Hansen[a] and Maria Andrée[b]

[a]Swedish National Agency for Education, Stockholm, Sweden; [b]Department of Mathematics and Science Education, Stockholm University, Stockholm, Sweden

ABSTRACT

Background: Research on peer assessment has noted ambiguity among students in using peer assessment for improving their work. Previous research has explained this in terms of deficits in the student feedback, or differences in student views of what counts as high-quality work.

Purpose: This study frames peer assessment as a social process in the science classroom. The aim is to explore peer assessment in science education as social practice in order to contribute to an understanding of the affordances and constraints of using peer assessment as a learning tool in science education.

Design and Method: The study was conducted in four lower secondary school classes, school years 8 and 9, in two different schools. An intervention study was designed focussing on the topic of experimental design. It involved the students in a process of peer assessment where they designed experiments individually, and then exchanged their designs, conducted each other's experiments, provided feedback to each other and revised their original design after discussing the feedback in groups. Data were collected in the form of audio recordings of student discussions and written work.

Results: The results show that, although not all peer feedback resulted in revisions, peer feedback was useful to the students in group interaction when negotiating quality in their work.

Conclusions: To conclude, the potential for using peer assessment in science education should not only be evaluated through the students' revisions but also in terms of in what ways the feedback constitutes interactional resources for defining quality in student work.

Introduction

Peer assessment (PA) generally refers to 'a process whereby students evaluate, or are evaluated by, their peers' (van Zundert, Sluijsmans, and van Merriënboer 2010). PA has been used for both formative and summative purposes (Falchikov 1995; Topping 1998). Summative uses may be motivated by aims to save teacher workload (Sadler and Good

2006). However, it may also be used as an assessment tool for teachers to discern the performance of individual students working in group projects (Cheng and Warren 2000). Sometimes PA is used formatively to complement teacher feedback because it is argued that feedback from peers may be easier to understand than feedback received from a teacher (Black and Wiliam 1998). Another use of PA is as a learning tool to develop students' abilities to form judgements about what constitutes high-quality work (Kollar and Fisher 2010; Strijbos et al. 2009; Topping 1998). Panadero (2016) argues that there are two potential outcomes of PA: improving work from feedback and learning from assessing. The outcomes of both depend highly on the framed purpose of PA (Harris and Brown 2013) and what guidance, for example, in the form of rubrics, are available to the students (Panadero, Romero, and Strijbos 2013). The study presented here concerns PA as a learning tool and the affordances and constraints for using PA as a learning tool in science education.

Previous research on peer assessment in science education

In our review of previous research on PA we have discerned three major challenges concerning peer assessment as a learning tool of relevance across national and institutional boundaries: (1) discrepancies between peer and teacher assessment, (2) how students use peer feedback and (3) the influence of human and social variables on processes of giving and receiving feedback.

(1) *The challenge of discrepancies between peer and teacher assessment*. Studies of PA at the university level show that university students commonly score and value the work of peers somewhat differently than their assessing teacher (Poon et al. 2009; Tal 2005). Another study by Hovardas, Tsivitanidou, and Zacharia (2014) found that university science students tended to emphasise content before skills to a larger extent than the teacher and that the students were less likely to find scientific mistakes than the teacher. Differences between the assessment of teachers and that of peers may be influenced by the task to be assessed, the criteria and experience of the assessor. However, Tsai, Lin, and Yuan (2002) found the differences between students' and teachers' assessment to be related to how they judged the relevance of the criteria at hand. Although the above research on discrepancies between peer and teacher feedback concern university level education, the recognition of potential discrepancies also in lower levels of education could be relevant in spite of differences in educational practices otherwise. For example, for students in science education, the consequences of discrepancies between teachers' and students' sense of quality may be large in relation to summative assessment – where peer assessment may impact grades. The challenge of discrepancies between peer and teacher assessment points to the importance of developing a shared classroom practice of what counts as *good quality* of work in assessment (Sadler 2007).

(2) There is tendency that *students do not fully utilise the peer feedback* they receive to improve their work (Jönsson 2013; Tsai, Lin, and Yuan 2002). One reason suggested is that the feedback is perhaps not helpful enough (Jönsson 2013). However, Tsivitanidou, Zacharia, and Hovardas (2011) noticed that upper

secondary school science students did not improve their work as suggested in the peer feedback, even when the feedback, from the perspective of the researchers, offered appropriate suggestions for improvement. Thus, students' reluctance, or failure, to use the received peer feedback cannot only be understood in terms of low quality of the feedback per se (Jönsson 2013). Jönsson (2013) has suggested that students might lack strategies for how to deal with the feedback they receive, or that students may perceive it unnecessary to utilise the feedback (as they may not actually be required to improve the work that has been peer-assessed). Another explanation, suggested by Tsivitanidou, Zacharia, and Hovardas (2011), is that there may be discrepancies between how different students understand assessment criteria. Therefore, the students may not perceive the feedback as relevant. This second challenge points to the importance of developing a shared practice of what counts as *useful* feedback in assessment (Sadler 2007)

(3) The *processes of giving and receiving feedback* are *influenced by human and social variables*. Human and social variables relevant to PA processes include prior experiences of PA, motivation for engaging, trust in both oneself and the peers as assessors, as well as comfort in the sincerity and appropriateness of feedback (Panadero 2016). Harris and Brown (2013) found in their multi-case study that students were well aware of the public nature and accountability aspect of the assessment; they tended to write superficial comments rather than articulating strengths and weaknesses. Moreover, the students did not credit their peers with the same trust as they did their teacher. However, Cheng and Tsai (2012) have argued that trust in peers could increase through improved PA practices. Gamlem and Smith (2013), who interviewed teenage students about how they generally perceived classroom feedback, found that the students wanted more specific suggestions on how they could improve their work. Also, the students questioned the, often compulsory, element of appraisal in feedback given to peers, since they sometimes found it difficult to find relevant aspects to praise. Consequently, they did not trust the honesty of the appraisals received. Somewhat contrary to the findings of Gamlem and Smith are the findings of Brown and Glover (2006), who found that students in an interview setting often expressed appreciation of peer feedback. However, they seldom used peer feedback to improve their work. In addition, Tsai, Lin, and Yuan (2002) found that social relationships between students, such as friendship status or position as a 'good achiever', influenced how feedback was provided between students. They even found examples where students withheld feedback for fear of letting peers outshine them. On the other hand, Panadero, Romero, and Strijbos (2013) found that the students in their study valued the quality of peers' work highly, regardless of friendship status. PA among friends could, however, be perceived as more comfortable (Harris and Brown 2013). The third challenge points to the importance of developing teachers' awareness of *conditions that influence the functionality of assessment* to improve the quality of work (Sadler 2007).

Based on a literature review on PA, van Zundert, Sluijsmans, and van Merriënboer (2010) points to the limitations of previous research in that it commonly focussed on single aspects of PA mostly within higher education. They argue for the need of research

to study the effects of both giving and receiving feedback, studying other fields than tertiary education, as well as studying content-specific PA. To date, research on formative PA has primarily focussed technologically mediated PA on tertiary education, for example, tasks of online comments on written productions (Huann-Shyang et al. 2011; Nicol 2009).There is thus a need for research at different levels of education that examines all three challenges of PA concerning how to develop a shared sense of disciplinary quality of work as well as quality of feedback and use of feedback to improve work in PA. A deeper understanding of how the three challenges and aspects of quality in different levels of education would be foundational for creating conditions in classroom practice where PA may contribute to student learning.

The aim of this study is to explore PA practices in lower secondary science education in order to contribute to an understanding of the affordances and constraints of using PA as a learning tool in science education. This aim is in line with the shift suggested by Jönsson (2013) concerning how to conceptualise peer feedback, from a transmission model of feedback (where the student is positioned as a receiver), to a more dialogic view of giving and receiving feedback as embedded in classroom practices. The research question is:

- How do students negotiate whether or not, and then how, to use peer feedback to improve their work?

Peer assessment as participation in peer assessment practice

In this study, we draw upon the theory of communities of practice (CoP) as developed by Wenger (1998) in order to theorise how students use peer feedback in a science classroom practice. This means that participation in PA is conceptualised as a process where students 'play the game' of PA (Willis 2011). From a CoP perspective, in PA, students negotiate quality both in the positions as givers and receivers of feedback. A presupposition is that student participation in PA is not framed as individual traits, but as socially negotiated ways of 'how to play the game' within a specific educational practice.

The theory of CoP provides some central conceptual tools for analysing social practices including community, participation, repertoires and reification. According to Wenger (1998), the notion of community refers to how groups form through mutual engagement in joint enterprise, and in developing shared repertoires for participation. This study analyses the participation of students in terms of how they engage in the science classroom practice: the extent to which they engage in a joint enterprise involving PA and develop shared ways of approaching the PA task.

Learning in a community of practice involves processes of participation and reification (Wenger 1998). Participation in a PA task refers to a process of participating in and sharing experiences with others, as well as negotiating what matters. What should be valued as good quality in the particular classroom? Reification involves processes such as 'making, designing, representing, naming, encoding' as well as 'perceiving, using etc.' (Wenger 1998).

Rubrics and scoring guides are formal examples of reification of an assessment practice. In communities of practice, formal reifications work as reference points. The

processes of making knowledge and learning processes explicit as part of PA requires relations between the explicit and tacit to be negotiated. In the same way, a reification – such as the formalisation of knowledge in a rubric – cannot simply be translated into participation but require individual and collective renegotiation. Thus, stressing formalism by introducing formal rubrics into classroom practice without matching participation might result in the rubrics being perceived as obsolete. In contexts where rubrics are externally determined – as is the case in Sweden where relatively generic rubrics are determined by the Swedish National Agency for Education – it becomes important to scrutinise student meaning-making of rubrics in PA; How do students transform what has been made explicit in the rubrics in PA tasks when articulating quality in the work of a peer? According to Wenger, 'what is said, represented, or otherwise brought into focus always assumes a history of participation as a context for its interpretation' (Wenger 1998, p 67). When students are afforded opportunities to communicate experiences of assessment, or assess together, this may contribute to the negotiation of a shared assessment repertoire and to 'converge' student participation in the assessment practice (Wenger 1998).

The limitations of PA identified in previous research concerning students' hesitancy to use peer feedback to improve their work, can be hypothesised to be related to lack of a negotiated assessment repertoire and a shared reified sense of 'good quality' and what it takes to be a 'good achiever'. For educational purposes, participating in PA discussions offers the potential to substantialise abstract criteria and reify a joint sense of quality both through the improvement of the students' work and reflections on what counts as high quality in the science classroom (Anker-Hansen and Andrée 2015; Dixon, Hawe, and Parr 2011; Kuipers 2011). One can also argue, in line with Heritage and Wylie (2018), for the importance of recognising and scrutinising the norms of participation in assessment for learning practices, by detailed analysis of classroom practices in relation to educational goals of equity and developing student learner identities.

Methods

Design and procedures

In order to study the processes of how students make use of and reject peer feedback in the science classroom, we set up an interventionist study. The intervention consisted of a lesson sequence about scientific inquiry (SI), where PA was embedded in the form of peer review processes and critical examination of the experimental design.

The rationale for organising PA in the form of peer review is that the latter is a central aspect of scientific practice and could lead to the development of students' ability to critically examine scientific processes (Kolsto 2001; Murphy, Lunn, and Jones 2006; Sandoval and Reiser 2004). Commonly, scientists worldwide conduct research that is very similar to and relates to the results of peers (Timmer 2012). Previous research has pointed to the importance of teaching aspects of SI explicitly to develop students' abilities to critically examine scientific work (Lederman, Lederman, and Antink 2013). One way of providing students with opportunities to experience critical examination of scientific processes in science education may therefore be to engage students in PA of experimental work. The objective to critically examine scientific work

has been part of the Swedish national curricula for a decade. In the latest curricula from 2011, the development of students' abilities to design, conduct and evaluate systematic investigations is formulated as an overall objective in the compulsory school (The Swedish National Agency for Education 2011). The topic of designing and evaluating systematic investigations is thus well suited for embedding PA as part of a lesson sequence.

A general design principle, that became important in relation to the choice of content, was to open up opportunities for students to expand and draw on personal funds of knowledge (Andrée and Lager-Nyqvist 2012; Barton and Tan 2009). By valuing and legitimising students' funds of knowledge as related to and applicable to work in the science classroom, we wanted to create conditions to afford students' participation and engagement in classroom work. Therefore, it was decided to focus the lesson sequence on the topic of food, nutrients and health, as this is a topic expected to be perceived as personally meaningful by the students (Andrée and Lager-Nyqvist 2012; Arvola and Lundegård 2012; Barton and Tan 2009; Jidesjö et al. 2012).

The explicit aspects of SI in focus in the lesson sequence were (A) avoiding confirmation bias and (B) controlling variables. We drew upon a model of designing scientific experiments used in the national assessments of the biology subject for school year nine, provided by The Swedish National Agency for Education (2009). This model (see Table 1) was thus familiar to teachers and students in Sweden; they were expected to use it in the national test on science in school year nine. Since experimental design items in the Swedish national tests, at the time, were rather open-ended, a variety in students' experimental designs was always expected, and the quality was not explicitly articulated in the scoring guides. However, a criterion was that the design contained reflections on sources of error and explicit instructions on how to deal with these when carrying out the experiment.

The focus of the preparations for the intervention had been on teaching the quality of SI. The students had been given lessons where the nature of SI was taught explicitly. In relation to PA, however, the students had no experience from prior science education, and the students were given the same assessment instructions available to teachers. In these instructions, quality was described generally as how sufficient the design was to produce a valid result or if it needed adjustments to be understood or to gain better control of variables. Panadero, Romero, and Strijbos (2013) have shown that scaffolding students with explicit rubrics could improve the precision of PA.

The students were assigned the task of designing an experiment individually, where they compared the effect of two different breakfasts on a certain form of physical morning exercise. In order to afford contradictions concerning SI, the task was intentionally constructed to allow for irremediable obstacles concerning variable control and confirmation bias. The choice of breakfasts and the physical exercises were, thus, left to the designing students. The experimental design produced by each student was then

Table 1. Model for students' experimental design.

- Description of equipment needed
- Description of how the experiment should be conducted
- Explanation why the experiment should be conducted this way
- Description of sources of error and how to deal with these
- Description of safety risks and how to deal with these

given to another student with an assignment to conduct the experiment (to eat the two different breakfasts exactly one week apart and measure the effect by means of the prescribed physical exercise). After they had conducted the experiment, the students were told to write individual feedback to their peers concerning what difficulties they had experienced, what results they had achieved, and what they suggested to improve the validity of the experiment. This review process thus differed from scientific reviews, where the reviewers do not actually conduct the reviewed study. However, we estimated that having concrete experience of the experiment would afford reifying the quality of the design in the feedback.

The teachers administered the exchange of the students' designs. The teachers chose to assign reviewers between students with approximately equal grades in the science subjects. The rationale was that it may be difficult for students with low grades to discern issues in need of improvement when reviewing the work of a high-achieving student. Although previous research points to anonymity in PA as a means to make the process more comfortable (Cheng and Tsai 2012), it was not possible to make the review anonymous since the students were going to conduct the experiment designed by a peer during class in the school facilities. Given the research of Gamlem and Smith (2013), where students felt the need to make things up in the feedback, no requirements were set regarding an upper or lower limit for how many words the students should write, nor were there any demands of including appraisal elements. Instead, the students were asked to write as much feedback as they felt necessary to improve the validity and reliability of the design.

In the final lesson, the review comments were handed back to the student who had originally designed the experiment and the students were asked to form groups of 3 to 7 to discuss the revisions of their experiments. The group discussion was intended to provide opportunities to study how the students negotiated using the received peer feedback with other students with whom they felt comfortable to work (Harris and Brown 2013).

Study setting

The lesson sequence was planned in collaboration with two science teachers at two different lower secondary schools in Sweden. The two teachers each had approximately 10 years of science teaching experience. The lesson sequence was staged in one class from school year eight (17 students, approximately 14–15 years old) and in three classes from school year nine (24 students in class A, 29 students in class B and 28 students in class C, approximately 15–16 years old). We secured written informed consent from all students and their caretakers for participation in the study. The peer review task was framed as a part of teaching SI practises. In addition, neither the experimental designs nor the peer feedback were used as a basis for teacher summative assessment and thus not included in the teachers' grading material.

Data collection

Data was collected in the form of audio and video recordings as well as the collection of copies of the texts produced by students during the four lessons in the four classes. The collected student work included the experimental designs, feedback and subsequent alterations of the original designs (see Figure 1). During the final lesson, where students
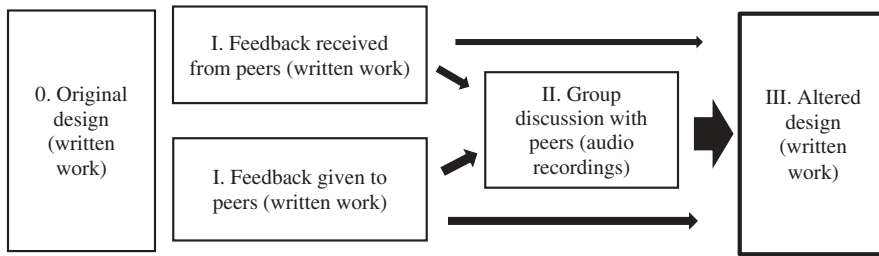
Figure 1. Data sources for how students negotiated how to use given and received peer feedback.

discussed PA and alterations, audio recorders were placed on the tables where the groups were sitting. Additionally, two opposing video cameras were used in the classroom to support transcription of the audio recordings.

Data analysis

The analysis of data was conducted in two steps. First, we analysed the written texts produced by each student to discern the alterations and source of alterations made to the students' designs (see 0, I and III in Figure 1). Second, we analysed the group discussions to discern of episodes when the students negotiated in groups how to use feedback to improve their work (see II in Figure 1). We specifically studied how quality as reified in different feedback was discussed in the groups and how the students negotiated what revisions to make in their experimental designs based on the peer feedback they received.

Analysis step one: student texts

In the text analysis we focussed on tracking changes from each student's text documents of the original design (0), through both given and received feedback (I), to the revised design (III). The changes were transcribed into digital text documents and linked with transcribed audio recordings through the NVivo 10 software (see Figure 1). Thus, it became feasible to follow what revisions the students discussed, (II) and what resources they used when revising their design (III). If, for example, a student chose to change the physical exercise in his or her experiment, for example, from counting push-ups to running a specified distance, we could determine if this was attributed to peer feedback the student had received in writing or group discussions, or if the changes had been suggested by the student him or herself in the feedback given to another student. However, it proved too difficult to divide feedback suggestions into quantifiable fractions. Thus, we chose to study only if the students had used some of the suggestions in the feedback, but not to *what extent* they had followed everything that was suggested.

Analysis step two: group discussions

To further understand how the students had made the decision to alter their original design, we focussed on how the students negotiated what revisions to make in their own

experimental design with other students in the group discussions (see II in Figure 1). We analysed the student utterances in the discussions regarding what experiences they mentioned as relevant for improving their work using the CoP concepts of participation and reification to analyse how the experiences were negotiated in conversation. Student participation was interpreted concerning what position and roles the students took, for example, if they spoke about themselves as researchers, consumers of research, or students. Reification was interpreted as articulation of what counted as quality. If the students, for instance, suggested alterations to make the experiment confirm what they already knew about healthy food, the students reified the confirmation of prior scientific knowledge as an aspect of the quality of an experiment.

Results

Revisions of student texts

Out of the students participating in the intervention, approximately four-fifth of the students chose to make some revision (see Table 2). Half of the students decided to make alterations that had been suggested in the received peer feedback and about as many made alterations that they had suggested themselves in feedback to a peer. Additionally, a small number of students made revisions of the final text based on the small group discussions.

Group discussions

In the group session, the students discussed how to handle the received peer feedback in preparing their final experimental design. For the most part students gave examples to each other of feedback that they had received, and the group then negotiated how to use the feedback.

The given and received peer feedback included suggestions concerning personal experiences and preferences (eg 'I exchanged the peanut butter with marmalade, because I do not like peanut butter.'), the experimental design (eg 'You could have specified how many sandwiches you were supposed to eat or whether it should be white bread or wholemeal.), and alterations to make the experiment fit a particular hypothesis better (eg theoretical knowledge should be used (eg 'You could choose two breakfasts that both bring satisfaction, where breakfast 1 does not contain much energy that lasts long and breakfast 2 contains more longer-lasting calories.'). However, there are also examples of non-specific feedback without suggestions for alterations, for example general appraisal.

Table 2. Model for students' experimental design.

Students who decided to make alterations from the first to a second version	79%
Students who decided to make alterations from feedback they had received from peers	52%
Students who decided to make alterations from the feedback they had given to peers	54%
Students who decided to make alterations from suggestions only mentioned in the group discussions	10%

We discerned three types of trajectories when students raised concerns about the different types of feedback in their groups:

- The group supporting a decision to disregard feedback
- The group emphasising the use of feedback
- The group redefining the quality of what constitutes a 'good experimental design'

To illustrate the processes of negotiation of the feedback received and quality in experimental design we present the three types of trajectories in the form of three episodes following the paths of three students: Tomas, Patricia and Carin. The episodes were chosen to illustrate the different directions and consequences of the students' small-group negotiations concerning how to deal with the received peer feedback: from disregarding feedback (Tomas), to emphasising the use of feedback (Patricia), or redefining the quality of what constitutes a 'good experimental design' (Carin). The discussion displayed in Tomas' and Patricia's groups represent negotiations that each could be found, to some extent, in more than half of the groups, while the discussion of Carin's group was unique to that group. The presentation of the results is intended to provide an account of how given feedback related to received feedback as well as the different ways in which the groups of students negotiated whether to reject or accept the written peer feedback. Since the experimental designs of the students were extensive, we present a summarised description of the original design and what was altered.

Negotiating how to disregard feedback following the path of tomas

Tomas had planned an experiment comparing a small and large breakfast by means of physical exercise through measuring the time it took to do a combination of 25 push-ups, sit-ups and high jumps. Tomas received the following peer feedback report:

Breakfast 1: I ate the banana. Instead of a boiled egg, I made scrambled eggs. Then I drank the juice. I didn't have the strength to do so many push-ups, so I counted sit-ups. My activity was sit-ups in 1 minute and 30 seconds. I managed 15.

Breakfast 2: I could only manage 2 sandwiches and 1 banana plus juice. I could only do it for 44 seconds.

Sources of error before the investigation: Sleep the same amount both days so that you have the same condition.

Sources of error after the investigation: Not so large breakfast no 2. It included 4 sandwiches, 2 bananas and a glass of juice because not everyone can eat that much. It is nearly the same. Was no 2 the healthy one? I got a worse result there. It would have been better if you showed which one was healthy and which one was unhealthy.

The above feedback is presented as changes already made when conducting the experiment. The reviewing student changed the second breakfast and altered the physical exercise. The feedback furthermore concerned the object of study, where the peer

questioned what was to be tested in the experiment and suggested a change to two more obviously healthy/unhealthy breakfasts. The report also made an alteration of the experiment from one boiled egg to scrambled eggs in the first breakfast, which was not explained. 'Sources of error before the investigation' referred to what Tomas had already written in his design.

The feedback Tomas gave another student included feedback concerning variable control by specifying the size of the suggested breakfasts: 'I would like to know how many sandwiches I was supposed to eat for breakfast 2' […] 'It would be better to know the number'. The rest of the feedback only concerned appraisals: 'It was tough. I got a sore throat, but it went well otherwise' […] 'I think it was good. Moderately tough.' The received feedback thus diverged from the feedback Thomas had provided in multiple ways. Different aspects of quality were reified in the received and given feedback. Whereas the received feedback concerned adjustments to suit the individual capabilities of the student conducting the experiment, the given feedback concerned the robustness of the experiment. Furthermore, the student reviewing the experiment design had understood the experiment to be a comparison between a healthy and an unhealthy breakfast, whereas Thomas had stated it was a comparison between breakfasts of different sizes. Finally, the feedback given by Thomas contained appraisals, whereas the received feedback did not.

In the group discussion on the received feedback, the group in which Tomas participated spoke about the purpose of the PA regarding appraisal.

1. Malin: Well, how did you all do?
2. Tomas: Mine went awry. She didn't understand anything. She thought it was lousy. She's, like, I don't understand anything. She was just negative and she thought it was too much with four sandwiches. She couldn't manage.
3. Annette: I think mine went really well. Mimmi did exactly, and she wrote: Really good, like I can understand. So it was great.
4. Tomas: [laughter] That was awesome!
5. Annette: She was really pleased with my inquiry.
6. Malin: What alterations did she suggest?
7. Annette: Eh, or, like, she has not written that, but I have seen things that I should improve.

In the above interaction Tomas reified the negative feedback where his peer had been unable to carry out the experiment, participating as a criticised and misunderstood designer (utterance 2). Annette contrariwise participated as a satisfied designer, reifying appraisals (3 & 5), which were supported by Tomas (4). When Malin asked about what suggested alterations Annette had received (6), Annette mumbled and answered that 'she has not written that' (7). The feedback reply reveals that Tomas received suggestions on how to change his design, but he expressed disappointment about the report. Both Annette and Tomas thereby reified a purpose of PA as receiving approval from their peers (rather than receiving useful advice on what revisions to make). The student group then continued to reify much of the feedback Tomas had received as irrelevant.'

8. Tomas: I can tell you how it went. She writes like this. I can tell you what she writes. She writes: To improve your inquiry, you would have to plan for another breakfast, because you should inquire. It would be better if you, if I, if you could conduct another type of activity. Perhaps she added this when she was done. She had to struggle a bit too long. Then she complains that there were too many sandwiches and that it was sandwiches.
9. Annette: Were the sandwiches big?
10. Tomas: But the second breakfast had the best results.
11. Annette: Mmm.
12. Tomas   And do two bananas. It was rather much [laughter].
13. Robert: But what the heck. You have to manage some.
14. Annette: [laughter] I can't understand what she couldn't understand. It's pretty obvious.
15. Tomas: Neither do I. And she mentions this with sleep as well. And I wrote this as a source of error. Darned if I'll write that again.
16. Annette: But you'll have to because the teachers will read your own.
17. Tomas: She wants this one to be an unhealthy breakfast and that one a healthy, but I thought it would be a small breakfast and a mega breakfast, sort of. So I went for it.
18. Robert: Were four sandwiches a small breakfast?
19. Tomas: Four sandwiches, two bananas and a large glass of juice.
20. Robert: Well, and what was the other one?
21. Tomas: It was only a banana, a glass of juice and a boiled egg. But she scrambled the eggs instead.
22. Robert: [laughter] That's a bit more than one egg.
23. Tomas: She can put whatever she wants in her mouth. No, I'm disappointed.

When comparing Tomas' utterance on the reviewer feedback (utterance 8) with the written reviewer feedback (see above), we found that Tomas was not quoting the peer review verbatim but rather expressing an interpretation. Annette and Robert confirmed Tomas' expressed disappointment as being legitimate with expressions such as 'But what the heck' and 'I can't understand what she couldn't understand' (see 13 & 14). Tomas was participating as a misunderstood designer, expressing reluctance to address the suggested alterations (15 & 17). Annette reminds him that the PA is a school task and that the teacher will expect him to make revisions (16). In the discussion between Tomas and the assessing student, Robert participated as mediator by inquiring whether there was a call for objection to the large breakfast (18 & 20). The episode is closed with Tomas expressing disappointment (23).

In the final design, Tomas wrote that it might be necessary to review the second breakfast and reduce the number of sandwiches. However, he did not actually change the experimental design.

Negotiating how to use feedback following the path of Patricia

Patricia had planned an experiment comparing a 'healthy' breakfast (two slices of whole-wheat bread with butter, tomato or cucumber and ham or cheese) with an 'unhealthy'

breakfast (two slices of toasted white bread with butter and cacao hazelnut cream or instant chocolate and one glass of a soft drink). The physical activity planned to evaluate the breakfasts included running as many laps on a track as one had the strength to in seven minutes and thereafter completing as many high jumps as possible in three minutes. Patricia received the following peer feedback:

> Whoever wrote this should only take one activity, because two take too much time. It was not tough, but one activity is enough for the investigation. There was nothing about what I was supposed to drink in the first task. Then you did not give any specific time when to go to bed. It is easier if you have a certain time instead of remembering when you went to bed the first time [...] We do not run very long and only eat two breakfasts, so the result would be better if you did it during a longer time. If you are going to do this task again, you need the same status in both tests (except breakfasts).

The above report includes suggestions of variable control. The suggestion to specify bed time was warranted with what is easier to remember. The suggestion to reduce one of the physical exercises was explicitly not attributed to personal experiences of fatigue, but rather what was sufficient for the investigation during the time available. Additionally, suggestions on robustness of the experiment further concerned desires to expand the experiment over a longer time period, specifying beverage and controlling variables (status) beside breakfast.

The feedback Patricia gave to another student similarly included suggestions on variable control, where she expressed: 'it was influenced by how much you had slept or if you had exercised before one time but not the other'. She also wrote: 'It would have been better to run during a specified time instead of running a number of laps'. This suggestion was unwarranted in the written feedback. However, during the class she talked about having 'made an extra effort' while running after the 'healthy breakfast', indicating addressing the problematics of confirmation bias.

In the group in which Patricia participated, the students talked about the PA process as providing opportunities for learning from others.

24. Patricia: Shall we discuss this? What do you think about reviewing each other's experiments like this?
25. Eva: I think it's really good. Perhaps this person did like this and it wasn't so successful, so you can compare with each other and make an improvement.
26. Mikaela: You learn from each other.
27. Patricia: Yes, exactly. You learn from each other.
28. Eva: Yes. And you see how other people think when they make evaluations and then you can find the best way to think and do more.
29. Patricia: Yes. Because everybody has different experiences and then you learn that this was better than this one.

Eva reified the benefits of comparing results and designs for improving their own experimental work (utterances 25 & 28). Mikaela and Patricia confirmed that you 'learn from each other' (26 & 27). Thus, before opening a conversation on the specific reviews received, the group collectively reified PA as an opportunity for improving their own work.

In the subsequent discussion, Patricia expressed difficulties concerning how to address the review comments about specifying bedtime:

30. Patricia:		Then he only wrote: You didn't g ive me any specific bedtime. I can't very well force you to sleep. You can go to bed whenever you like.
31. Mikaela:		[laughter]
32. Martina:		But if this was a big inquiry and 110% serious, you don't write that you should go to bed like eight. But if you like. I wrote like this. If you really want a serious result, which is really clear, then you should tell the person when to go to bed.
33. Eva:		Yes. Then everyone has to go to bed at, like, eleven. Then we all have to sleep as much.
34. Patricia:		But, it's hard to just go.
35. Martina:		Yes, but if a professor did it.
36. Patricia:		I think it's enough if you go to bed the same time as last time.
37. Mikaela:		I wrote that you should sleep seven hours so that you sleep well and can manage more.
38. Patricia:		Yes, but I don't know if it becomes much better.

Patricia referred to the peer's suggestions on a specified bedtime and objected that 'I can't very well force you to sleep' (utterance 30). After initial confirming laughter (31), Martina introduced a 'but if' asking Patricia to imagine the experiment as 'serious' and conducted by a 'professor' (32 & 35). The imagination is elaborated and supported by Eva and Mikaela (33 & 37). Although Patricia initially sustained her position of not wanting to impose control (34), she did adapt what the group reified as a more serious approach to research (36). However, she remained sceptical of whether further specification would matter in the end (38).

In the final report Patricia replied to feedback received by changing instructions, from stating that rest should be constant, to a clarification that the test person should go to bed at the same time. She also commented on the need for further rigour of the experimental design if the results of the experiments were to be published in a newspaper, and that a scientist should probably be more thorough in his or her design. In the group discussion, Patricia was encouraged to adopt a more scientific mode of participation by means of imagination, and thereby reifying variable control as an important aspect of the experimental design.

Negotiating How to Redefine the Quality of What Constitutes a 'Good Experimental Design' Following the Path of Carin

Carin had planned an experiment comparing a breakfast rich in dairy products and a breakfast rich in carbohydrates by means of a physical exercise of running as many laps with high knees along a soccer field in three minutes. Carin received the following feedback:

> *Breakfast: I did not eat anything for breakfast because I have been allergic to milk my entire life, and I do not like the taste because of that. Additionally, it is not healthy. I feel sick and get*

rashes. Then I thought I could compare not eating and then eating my sandwiches. I ran a lap on the white track because I cannot run with high knees for three minutes, and then I thought it would be better to do something that I can do completely. I thought Carin's investigation was good, except for the exercise because there will be many who cannot do it. I think the sources of errors are good and they are correct, and I will try to follow them.

Breakfast 2: I ate what the instructions said, except for just eating 2 slices of bread and ham, because 3 slices are a lot, I think, and 2 are enough for me.

Sources of error. If you don't sleep as much, it will influence the result. But it was hard to do something about that.

The above report includes suggestions of what was personally relevant to the student conducting the experiment. She had changed the experiment to comparing no breakfast with eating a breakfast of two slices of bread and ham. The reviewer wrote that she was allergic to milk and that she could not eat the first breakfast. She was participating as a person evaluating personal health and the experiment as a suggestion of what she (or lactose intolerant individuals) should eat for breakfast rather than a comparison between dairy products and carbohydrates, as Carin had stated. She further adjusted the physical exercise to something she thought she could manage.

The peer feedback Carin gave another student regarded variable control and confirmation bias: 'If I had been allowed to choose the exercise, I would not have chosen hover because it depends so much on willpower. I would have chosen some form of how many sit-ups I could have done during a minute or how fast I could run 100 m'. The qualities reified in the received feedback thus differed from the feedback Carin had given. Whereas the feedback received concerned the conditions of the reviewer, the given feedback concerned the robustness of the experiment itself.

In the group discussion, Carin brought up the question of whether a participant's inability to consume the suggested breakfasts should be considered a source of error or not:

39. Carin:	She has brought that up as a source of error. That it could have been vegetarian. And if you are vegetarian, it will go wrong. But she's not vegetarian, so…
40. 1st Author:	Is it a source of error to say that it's not vegetarian?
41. Amalia:	No.
42. Caroline:	No.
43. Carin:	I had it as a source of error in here, or something.
44. 1st Author:	Did you receive that?
45. Carin:	Yes. I have received that as an improvement.
46. Ida:	But, then you probably shouldn't have that as…
47. Caroline:	No. Not really.
48. Ida:	If you should do a real inquiry now, then it has to be people who can, like, eat the same thing. She couldn't run either.
49. 1st Author:	No?
50. Ida:	No. You have to have people who are somewhat similar, but yet different, or however you say it.

Carin told the group that she had received suggestions to include vegetarian alternatives (utterances 39, 43 & 45). When comparing the written feedback from the reviewer, one can see this was not correct. Prior to the interaction in the excerpt above, Carin had asked her group members: 'What do you say when milk or yoghurt does not contain lactose? Oat milk and there is soy yoghurt?' She received the replies 'lactose-free' and 'vegetarian' followed by a discussion on whether her peer reviewer really had dietary restrictions. The first author present did not understand what the discussion was about and asked the group if they considered 'not vegetarian' to be a source of error (40), to which Amalia and Caroline stated they did not. Ida suggested it would be better to find test people who could eat the specified breakfast and do the specified exercise (48 & 50). By doing this, Ida assumed the position of a scientist, defending the intention of Carin's study: to study a dairy-rich breakfast. The first author's questions may have caused the students to give the topic extra attention, but the group had not resolved whether to consider food allergy as a source of error:

51. Carin: But, then the fact that she's allergic is no longer a source of error.
52. Ida: But, source of error. Then I think, like this, variables that can influence the physical activity.
53. Jenny: But it can. If you are allergic to something, then…
54. Ida: Then she may have an allergic shock.
55. Carin: Yes.
56. Jenny: And then she cannot complete the design.
57. Carin: But then I'll take that. I'll take it.

In the above excerpt, Ida and Jenny elaborated on the definition of a source of error as variables that affect physical activities (utterances 52, 53, 54 & 56). The episode was closed by Carin accepting a food allergy as a legitimate source of error (57).

In the final report, Carin opens with the comment: 'I had forgotten to consider the risk that the test person could be allergic. In this case, to lactose'. Then she proposed specifying non-milk alternatives to the dairy products in the instructions. She also commented on the need for more participants in the study, over a sustained period of time, controlling for variables such as allergies or diseases that might affect the results. Thus, after negotiating how to interpret the written feedback in the group discussion, she did use the feedback to improve the design within the original rationale of the experiment.

Discussion and conclusions

The students in this study used both peer advice and their own experiences from evaluating the designs of their peers. Although students were sometimes initially reluctant to use feedback they received that differed from the feedback they had given, the quality and use of feedback was negotiated in the group discussions. Thus, the group discussions supported the students in interpreting and understanding the received written peer feedback.

Prior research suggests that one reason for rejecting PA could be that the peer feedback is lacking in quality. Possible reasons are that the PA offers unequal support

to students due to a variation of the quality of the feedback (Jönsson 2013), or that the students receiving the feedback may have unequal abilities to utilise the feedback in a revision of their work (Tsivitanidou, Zacharia, and Hovardas 2011). However, this study observed that the students' decision to reject feedback was not only a matter of how they initially valued the quality of the feedback, but also a consequence of the process of negotiating what was feasible and reasonable in the given teaching situation. In the group discussions, the students engaged in refining aspects of quality. However, this did not always result in an adjustment of the design. Consequently, we argue that the potential for PA in science class should not only be evaluated through the extent to which students perform revisions of a specific piece of work. Instead, potential benefits of peer feedback as interactional resources, affording reflections on the quality, should also be taken into account. Similar observations have been made in language education, where researchers have pointed to the importance of developing an awareness of strengths and shortcomings in peer texts for the further development of students' own texts (Lundstrom and Baker 2009; Min 2005). Lundstrom and Baker (2009) explain that when students read peer texts, they can search for solutions on how to overcome their own obstacles and avoid making the same mistakes as the reviewed author. As explained by Annette in Tomas' group, she could also make changes based on her own experiences of carrying out a peer experiment, even though she had not received any suggestions for improvements from her peer reviewer. Consequently, providing peer feedback could be useful for the reviewing peer, even though the specific suggestions received on their own work may not necessarily be useful.

Discrepancies between assessments performed by science teachers and students as well as between students are well known (Hovardas, Tsivitanidou, and Zacharia 2014; Poon et al. 2009; Tal 2005; Tsai, Lin, and Yuan 2002). Such discrepancies have been proposed as one reason for students rejecting peer feedback (Tsivitanidou, Zacharia, and Hovardas 2011). In this study, episodes of group interactions support this proposition. For instance, the discussion on Tomas' work is an example of how students supported each other in rejecting the suggestions based on personal experiences or questioning the object of study as misunderstandings. However, when listening to the student discussions, we also noticed that diverging views of quality are negotiated; such negotiations may open up for revisions, as in the discussions in Patricia's and Carin's groups. Harris and Brown (2013) showed that learning from PA was influenced by the framed purpose. We found that students also engaged in negotiating different reasons for giving and receiving feedback within a framed purpose. The students interviewed by Gamlem and Smith (2013) criticised compulsory components of appraisals; they desired more concrete suggestions on how to improve their work. However, in this study Tomas and Annette emphasised appraisals, and Tomas was reluctant to address the suggestions for change received. Conversely, Patricia's group had reified the purpose of PA as a task of exchanging experiences and improving the experimental designs. Patricia's friends also referred to this purpose when convincing her to address the suggested changes. Consequently, we suggest that another possible reason for whether students will make use of peer feedback relates to the extent a student shares the purpose of PA with his or her peers. This finding suggests that for PA to be effective, it is necessary that students share a purpose of assessment as a means to improve their work.

We agree with van Zundert, Sluijsmans, and van Merriënboer (2010) that there is a need for further research on how students' use of PA develops over time, and how students transfer the negotiated meanings of what counts as good quality longitudinally and across science contexts. In our study the students had not been provided any prior training or opportunities to develop their expertise in peer assessment. Nonetheless, students negotiated their decisions on how to use feedback when discussing this with peers. Thus, we hypothesise, in line with theories of communities of practice and the studies by Dixon, Hawe, and Parr (2011) and Willis (2011), that further engagement in the PA practices of experimental design will afford the development of joint assessment repertoires. To further develop PA practices, students likely need more teacher support and scaffolding concerning how to provide feedback compared to what was offered in this study (Harris and Brown 2013; Panadero, Romero, and Strijbos 2013). However, our analysis of student negotiations of the specific PA task in this study point to the importance of paying attention to how students experience aims of PA. Perhaps more importantly, providing students with opportunities to negotiate peer feedback in groups opens up for the collective negotiation of aims. It also has the potential to further develop students' awareness of the quality of their understanding of a specific content area (in our case, SI).

The intervention reported here offered students an opportunity not only to read and comment on the work of peers, but also to actually try the experimental designs practically. This is also an example of how science education offers specific conditions for PA relating to both classroom practices and to the objectives for student learning about SI. In other words, PA provides students with personal experiences of central aspects of SI such as a critical examination of scientific processes.

Disclosure statement

References

Andrée, M., and L. Lager-Nyqvist. 2012. "What do you know about fat? Drawing on diverse funds of knowledge in inquiry based science education." *Nordic Studies in Science Education* 8 (2): 178–193. doi:10.5617/nordina.526.

Anker-Hansen, J., and M. Andrée. 2015. "Affordances and Constraints of Using the Socio-political Debate for Authentic Summative Assessment." *International Journal of Science Education* 37 (15): 2577–2596. doi:10.1080/09500693.2015.1087068.

Arvola, A. O., and I. Lundegård. 2012. "It's Her Body. When Students' Argumentation Shows Displacement of Content in a Science Classroom." *Research in Science Education* 42 (6): 1121–1145. doi:10.1007/s11165-011-9237-2.

Barton, A. C., and E. Tan. 2009. "Funds of Knowledge and Discourses and Hybrid Space." *Journal of Research in Science Teaching* 46 (1): 50–73. doi:10.1002/tea.v46:1.

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5 (1): 7–74. doi:10.1080/0969595980050102.

Brown, E., and C. Glover. 2006. "Evaluating Written Feedback." In *Innovative Assessment in Higher Education*, edited by C. Bryan and K. Clegg, 81–91. New York, NY: Routledge.

Cheng, K. H., and C. C. Tsai. 2012. "Students' Interpersonal Perspectives On, Conceptions of and Approaches to Learning in Online Peer Assessment." *Australian Journal of Educational Technology* 28 (4): 599–618.

Cheng, W., and M. Warren. 2000. "Making a Difference: Using Peers to Assess Individual Students' Contributions to a Group Project." *Teaching in Higher Education* 5 (2): 243–255. doi:10.1080/135625100114885.

Dixon, H. R., E. Hawe, and J. Parr. 2011. "Enacting Assessment for Learning: The Beliefs Practice Nexus." *Assessment in Education: Principles, Policy & Practice* 18 (4): 365–379. doi:10.1080/0969594X.2010.526587.

Falchikov, N. 1995. "Peer Feedback Marking: Developing Peer Assessment." *Innovation in Education and Training International* 32 (2): 175–187. doi:10.1080/1355800950320212.

Gamlem, S. M., and K. Smith. 2013. "Student Perceptions of Classroom Feedback." *Assessment in Education: Principles, Policy & Practice* 20 (2): 1–20.

Harris, L. R., and G. T. L. Brown. 2013. "Opportunities and Obstacles to Consider When Using Peer- and Self-Assessment to Improve Student Learning: Case Studies into Teachers' Implementation." *Teaching and Teacher Education* 36: 101–111. doi:10.1016/j.tate.2013.07.008.

Heritage, M., and C. Wylie. 2018. "Reaping the Benefits of Assessment for Learning: Achievement, Identity, and Equity." *ZDM Mathematics Education* 50 (4): 729–741. doi:10.1007/s11858-018-0943-3.

Hovardas, T., O. E. Tsivitanidou, and Z. C. Zacharia. 2014. "Peer versus Expert Feedback: An Investigation of the Quality of Peer Feedback among Secondary School Students." *Computers & Education* 71: 133–152. doi:10.1016/j.compedu.2013.09.019.

Huann-Shyang, L., R. H. Zuway, W. Hsin-Hui, and L. Sung-Tao. 2011. "Using Reflective Peer Assessment to Promote Students' Conceptual Understanding through Asynchronous Discussions." *Journal of Educational Technology & Society* 14 (3): 178–189.

Jidesjö, A., M. Oscarsson, K. G. Karlsson, and H. Strömdahl. 2012. "Science for All or Science for Some: What Swedish Students Want to Learn about in Secondary Science and Technology and Their Opinions on Science Lessons." *Nordic Studies in Science Education* 5 (2): 213–229. doi:10.5617/nordina.352.

Jönsson, A. 2013. "Facilitating Productive Use of Feedback in Higher Education." *Active Learning in Higher Education* 14 (1): 63–76. doi:10.1177/1469787412467125.

Kollar, I., and F. Fisher. 2010. "Peer Assessment as Collaborative Learning: A Cognitive Perspective." *Learning and Instruction* 20 (4): 344–348. doi:10.1016/j.learninstruc.2009.08.005.

Kolsto, S. D. 2001. "To Trust or Not to Trust… Pupils' Ways of Judging Information Encountered in a Socio-Scientific Issue." *International Journal of Science Education* 23 (9): 877–901. doi:10.1080/09500690010016102.

Kuipers, J. 2011. "Unmentionable Others: Representing Participation Frameworks in School Science." *Anthropological Quarterly* 84 (1): 87–99. doi:10.1353/anq.2011.0002.

Lederman, N. G., J. S. Lederman, and A. Antink. 2013. "Nature of Science and Scientific Inquiry as Contexts for the Learning of Science and Achievement of Scientific Literacy." *International Journal of Education in Mathematics, Science and Technology* 1 (3): 138–147.

Lundstrom, K., and W. Baker. 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.

Min, H. T. 2005. "Training Students to Become Successful Peer Reviewers." *System* 33 (2): 293–308. doi:10.1016/j.system.2004.11.003.

Murphy, P., S. Lunn, and H. Jones. 2006. "The Impact of Authentic Learning on Students' Engagement with Physics." *Curriculum Journal* 17 (3): 229–246. doi:10.1080/09585170600909688.

Nicol, D. 2009. "Assessment for Learner Self-Regulation: Enhancing Achievement in the First Year Using Learning Technologies." *Assessment & Evaluation in Higher Education* 34 (3): 335–352. doi:10.1080/02602930802255139.

Panadero, E. 2016. "Is It Safe? Social, Interpersonal, and Human Effects of Peer Assessment: A Review and Future Directions." In *Handbook of Human and Social Conditions in Assessment*, edited by G. T. L. Brown and L. R. Harris. New York, NY: Routledge.

Panadero, E., M. Romero, and J. W. Strijbos. 2013. "The Impact of a Rubric and Friendship on Construct Validity of Peer Assessment, Perceived Fairness and Comfort, and Performance." *Studies in Educational Evaluation* 39 (4): 195–203. doi:10.1016/j.stueduc.2013.10.005.

Poon, W. Y., C. McNaught, P. Lam, and H. S. Kwan. 2009. "Improving Assessment Methods in University Science Education with Negotiated Self- and Peer-Assessment." *Assessment in Education: Principles, Policy & Practice* 16 (3): 331–346. doi:10.1080/09695940903319729.

Sadler, D. R. 2007. "Perils in the Meticulous Specification of Goals and Assessment Criteria." *Assessment in Education: Principles, Policy & Practice* 14 (3): 387–392. doi:10.1080/096959 40701592097.

Sadler, P. M., and E. Good. 2006. "The Impact of Self- and Peer-Grading on Student Learning." *Educational Assessment* 11 (1): 1–31. doi:10.1207/s15326977ea1101_1.

Sandoval, W. A., and B. J. Reiser. 2004. "Explanation-Driven Inquiry: Integrated Conceptual and Epistemic Scaffolds for Scientific Inquiry." *Science Education* 88 (3): 345–372. doi:10.1002/sce.10130.

Strijbos, J. W., T. A. Ochoa, D. M. A. Sluijsmans, M. S. R. Segers, and H. Tillema. 2009. "Fostering Interactivity through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments." In *Cognitive and Emotional Process in Web-Based Education: Integrating Human Factors and Personalization*, edited by C. Mourlas, N. Tsianos, and P. Germanakos, 375–395. Hersey, PA: IGI Global.

Tal, T. 2005. "Implementing Multiple Assessment Modes in an Interdisciplinary Environmental Education Course." *Environmental Education Research* 11 (5): 575–601. doi:10.1080/135046 20500169767.

The Swedish National Agency for Education. 2009. *Nationella ämnesprovet i biologi våren 2009* [The national subject test in biology spring 2009]. Sweden: Stockholm.

The Swedish National Agency for Education. 2011. *Curriculum for the Compulsory School, Preschool Class and Leisure-Time Centre 2011*. Sweden: Stockholm.

Timmer, J. 2012. *Scientific Reproducibility, for Fun and Profit. Arstechnica*. Accessed http://arstechnica.com/science/2012/08/scientific-reproducibility-for-fun-and-profit/website

Topping, K. 1998. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 68 (3): 249–276. doi:10.3102/00346543068003249.

Tsai, C. C., S. S. J. Lin, and S. M. Yuan. 2002. "Developing Science Activities through a Networked Peer Assessment System." *Computers & Education* 38 (1): 241–252. doi:10.1016/S0360-1315(01)00069-0.

Tsivitanidou, O. E., Z. C. Zacharia, and T. Hovardas. 2011. "Investigating Secondary School Students' Unmediated Peer Assessment Skills." *Learning and Instruction* 21 (4): 506–519. doi:10.1016/j.learninstruc.2010.08.002.

van Zundert, M., D. Sluijsmans, and J. van Merriënboer. 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.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. New York, NY: Cambridge University Press.

Willis, J. 2011. "Affiliation, Autonomy and Assessment for Learning." *Assessment in Education: Principles, Policy & Practice* 18 (4): 399–415. doi:10.1080/0969594X.2011.604305.