Predictive Effects of the Quality of Online Peer-Feedback Provided and Received on Primary School Students' Quality of Question-Generation

Author(s): Fu-Yun Yu and Chun-Ping Wu

Source: *Journal of Educational Technology & Society*, Vol. 19, No. 3 (July 2016), pp. 234-246

Published by: International Forum of Educational Technology & Society

Stable URL: https://www.jstor.org/stable/10.2307/jeductechsoci.19.3.234

Predictive Effects of the Quality of Online Peer-Feedback Provided and Received on Primary School Students' Quality of Question-Generation

Fu-Yun Yu[1] and Chun-Ping Wu[2*]

[1]Institute of Education, National Cheng Kung University, Taiwan // [2]Department of Educational Technology, TamKang University, Taiwan // fuyun.ncku@gmail.com // cpwu303@gmail.com

[*]Corresponding author

(Submitted February 11, 2015; Revised September 6, 2015; Accepted October 20, 2015)

ABSTRACT

The research objectives of this study were to examine the individual and combined predictive effects of the quality of online peer-feedback provided and received on primary school students' quality of question-generation. A correlational study was adopted, and performance data from 213 fifth-grade students engaged in online question-generation and peer assessment for six weeks were analysed using hierarchical multiple regression, with the dependent variable of scores on question-generation and independent variables of scores on peer-feedback provided and received. The results from the two-step hierarchical regression analysis indicated that the quality of peer-feedback provided and received, respectively, predicted students' quality of question-generation. Furthermore, the results from the three-step hierarchical regression analysis showed that the quality of peer-feedback provided and received in combination also predicted students' quality of question-generation. Details of the significance of this study are provided, as well as suggestions for instructional implementations.

Keywords

Peer-feedback provided, Peer-feedback received, Online learning activity, Peer-assessment, Question-generation

Introduction

Student question-generation as a contemporary approach to teaching and learning, and areas in need of future work

Student question-generation (SQG) engages students in reflecting on a recent learning experience, and constructing questions around areas they deem personally relevant and important for self- or peer-assessment. Its pedagogical significance and value have been well established empirically. Specifically, based on the results of a meta-analysis of 109 empirical studies on SQG conducted in a wide variety of disciplines and at all levels of schooling (with the primary level taking up nearly two-thirds of the studies), there is wide support for its positive effects on a variety of learning outcomes (e.g., academic achievement, attitudes toward learning) (Yu, 2012).

Despite SQG's solid empirical support, and sound theoretical foundations on the theories of self-regulated learning, constructivism, and self-determination, several factors affect its wider adoption in classrooms. In particular, studies show that primary school students have concerns about their capability and performance as related to SQG (Yu & Liu, 2005). There is thus a need to examine how to ease students' concerns and provide adequate scaffolding for SQG by taking advantage of peer-assessment (PA), as this approach not only allows students to receive more timely feedback in a large class, but also encourages them to keep examining the quality criteria of the expected learning outcomes (Topping & Ehly, 1998).

Peer-assessment as a generative learning approach and the current research gaps

The cognitive conflict, social constructivism, and social learning theories can help elucidate why the various cognitive processes brought about by PA (including self-evaluation, self-correction, self-adjustment, and self-reflection through giving, receiving and responding to comments) promote students' performance and cognitive growth (van Gennip, Segers & Tillema, 2010). Indeed, a growing body of empirical evidence since the 1990's has shown PA to promote students' critical thinking, cognitive development, and performance (Nelson & Schunn, 2009; Topping, 2010; van Gennip et al., 2010).

Efforts have been devoted to refine PA designs by identifying beneficial elements of PA tasks, such as the provision of clear and pre-specified criteria for objective comments and training (van Zundert, Sluijsmans, & van Merriënboer, 2010), the offering of both appropriate affective and cognitive feedback (Cho & Cho, 2011; Nelson & Schunn, 2009), and the inclusion of both quantitative and descriptive feedback (Topping, 2010; Yu & Wu, 2013). However, there are still a number of gaps in the related literature worthy of further investigation. In particular, few studies examine the effects of PA from both perspectives (i.e., the assessors and the assessed), or have clearly differentiated the effects of assessing peers from those of being assessed by peers (Topping, 2010; van Zundert et al., 2010). Moreover, those studies that have done so have all involved college or secondary school students. For instance, the results from Li, Liu and Steckelberg (2010), which involved undergraduate students engaged in computer-related projects, showed that the quality of comments given significantly predicted the performance of assessors, whereas the quality of received comments did not. The results from Cho and Cho (2011), which examined undergraduate students writing a science report on physics, similarly found that giving comments positively influenced the manner in which the assessors revised their own writing, although the effects of receiving peer comments were limited. Finally, Lu and Zhang (2012) and Lu and Law (2012) both examined secondary students and indicated that the cognitive feedback that was given strongly predicted how the assessors performed in their final projects, but was not related to the performance of the assessed student.

Topping (2010) noted that little research about PA was conducted in primary schools, and it was also suggested by van Zundert et al. (2010) that research on PA applied in contexts other than higher education should be a focus of future work, to extend the generalizability of PA. Since primary school students have the capability needed to engage in and benefit from PA (Hwang, Hung, & Chen, 2014; Yu, Liu, & Chan, 2005), by recognizing this gap in the literature the first focus of the current study is the respective effects of receiving and providing peer-feedback among primary school students.

Cognitive demands involved in providing and receiving feedback during peer-assessment: cognitive load perspective

It is generally believed that students can benefit from serving as both the assessor and the assessed in a PA context (Kollar & Fischer, 2010; Topping, 2010; van Gennip et al., 2010). On the one hand, the process of commenting on the strengths and weaknesses of assessed products may prompt the assessor to examine the evaluation criteria, and to develop knowledge of what constitutes good work and what needs to be avoided. This information can be further adopted by the assessor to monitor and regulate learning (Cho & Cho, 2011). On the other hand, those that are assessed can receive feedback from the assessor in a timely and personalized fashion, and make revisions to further enhance the quality of their work.

Despite the fact that simultaneously playing the roles of the feedback provider and receiver may yield some gains, playing both roles imposes cognitive loads on students. In short, while serving as the assessor, students have to study the assessed work first and then evaluate its quality against a set of criteria, followed by the construction of descriptive comments in a way that the assessed can sympathize with and benefit from. Similarly, when serving the role of the assessed, students have to process the comments provided, assess the validity and usefulness of each suggestion, and decide whether and how to integrate these to enhance their current work (Kollar & Fischer, 2010). As noted by cognitive load theorists (e.g., Paas, Renkl, & Sweller, 2004), any task composed of sub-tasks, each involving the activation and use of complex processes, demands cognitive capacity from the learner. If such tasks exceed the cognitive capacity of the learner, then this will cause a state of cognitive overload, which will lead to diminished learning effects.

Although there are studies that have students play only one role (e.g., Lundstrom & Baker, 2009), and others in which students play both roles, to the best of the authors' knowledge, no works have produced empirical data substantiating any added benefits from students playing both roles during PA. Therefore, the second focus of this study was to examine the combined predictive effect of both aspects involved in PA while playing both roles (i.e., peer-feedback received and provided) on the quality of students' question-generation. The combined effects of giving and receiving feedback on learning need to be substantiated to warrant the further use of such practices, especially for primary school students who, despite having reached Piaget's (1926) formal operational stage of being able to manage the logical use of symbols related to abstract concepts, are usually still very limited in term of cognitive capacity when compared to older students.

Research questions

The following two research questions are examined in this work:

- Will the quality of peer-feedback provided and received, respectively, significantly predict primary school students' quality of question-generation?
- Will the quality of peer-feedback provided and received, in combination, significantly predict primary school students' quality of question-generation?

Method

Participants

Two hundred and thirteen fifth-grade students (10-11 years old) from eight intact classes of one public primary school in Taiwan participated in this study. With the current emphasis on a competency-based curriculum targeting core competencies, including problem-solving, creativity, ICT skills, and independent thinking (Taiwan Elementary and Educator Community, 2014), the introduction of SQG and PA was thus well received by the participating school.

Research design and implementation procedures

The correlational research design method was adopted to guide the data collection and analysis. Hierarchical multiple regression was used to detect the effects of predictor variables on a criterion variable, with the criterion variable being the students' scores on the quality of question-generation, and predictors the scores on the quality of peer-feedback provided and received. Specifically, two-step hierarchical regression was employed to examine the predictive effect of peer-feedback provided and received, respectively, on the quality of question-generation, and three-step hierarchical regression to examine the predictive effect of peer-feedback provided and received, in combination, on the quality of question-generation, with significantly diminished learning effects connoting cognitive overload on the part of the learner, as suggested by cognitive load theory.

This study lasted for six consecutive weeks. Participants were informed that the online SQG and PA activity was intended to support their learning of two science units: (1) heat transfer and (2) sound and musical instruments. Since true/false and multiple-choice questions are among the most frequently encountered question types in primary schools in Taiwan, they were chosen for the focal online activities.

Each week, students headed to a computer lab to participate in a 40-minute online learning activity led by the implementer, after attending three instructional sessions on science taught by their respective science teachers in class. Training sessions that have been shown to be adequate for preparing participants of the same grade level to possess the fundamental skills required for question-generation and PA (Yu, Liu, & Chan, 2005) were scheduled in the first two weeks, before the subsequent four weeks of implementation. The experimental procedure is shown in Figure 1, and explained below.

A training session on generating the focal question types was arranged in the first week. Briefly, with reference to Rosenshine, Meister, and Chapman (1996), effective instructional elements supporting question-generation, namely the criteria and models of question-generation, were introduced (see "variables and measures" sub-section). In addition, based on the literature on test construction (Haladyna & Downing, 1989; Osterlind, 1998), basic principles for item writing and operational procedures for question-generation were illustrated in the Question-Authoring and Reasoning Knowledge System (QuARKS), followed by hands-on practice with the system.

At the beginning of the second session, whole-class feedback on students' question-generation performance was given to highlight good question-generation practices. PA training was then given, in which models of constructive feedback with reference to the devised scoring scheme (see "variables and measures" sub-section) and built-in criteria in the online PA form (see "online learning system" sub-section) were explained. Moreover, the operational procedures for PA and details of how to access peer-feedback in QuARKS were explained before the students practiced giving peer-feedback using the online PA form, and then receiving peer-feedback via the notification

system. The remaining time was devoted to individual hands-on practice of question-generation and PA using the system.

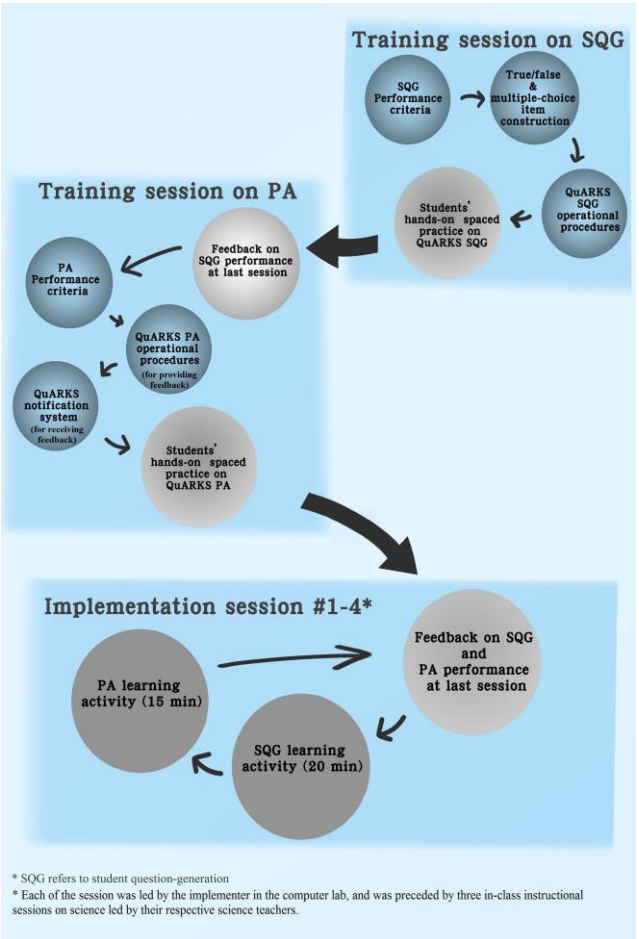

Figure 1. Experimental procedures

Starting from the third week, as a routine, whole-class feedback on good question-generation and PA practices was highlighted using the work of 3-5 students from the previous session. Afterwards, students were given 20 minutes to individually generate at least one question for each of the two chosen question types in accordance with the instructional content covered that week. Then, with the system's embedded automatic random assignment, students were directed to choose and assess at least two questions from a pool of eight peer-generated questions for each chosen question type for the remaining time (i.e., about 15 minutes).

Online learning system

The QuARKS online learning environment equipped with the usual advantages associated with computer and networked technologies (Yu, 2009) was adopted to support the learning activities carried out in this study. In brief, QuARKS allows students to contribute and benefit from the process of constructing question items of different types and media formats. Also included is a PA component to support question-generation activities. For PA, in particular, with reference to a set of built-in criteria (the bottom-left portion of Figure 2) and the scoring scheme for the quality of PA (see "variables and measures" sub-section for detail), the assessor first completes two rating scales with regard to: (1) the overall quality of the assessed question, and (2) the recommendation for the examined question to be included in the online drill-and-practice item bank (the top portion of Figure 2). The assessor then provides descriptive comments by typing in suggestions in a designated feedback space (the bottom-right portion of Figure 2).

To ensure the fluidity of the PA process, a notification system was put in place to automatically alert the question-author of any assessment updates. Specifically, an assessment is completed, the author of the question is alerted instantly by a blinking red icon on the screen. By clicking on the icon and the "proceed" button placed next to the notification, the author is transferred directly to the focal question, where its assessment can be viewed, and appropriate actions can be taken, such as making revisions.

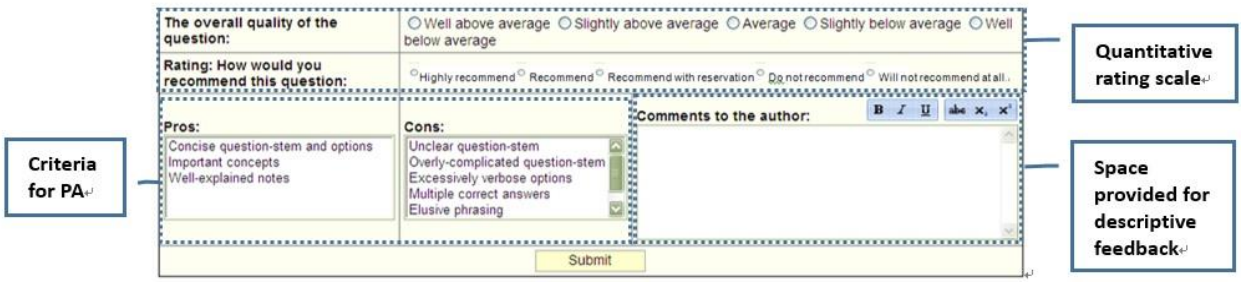

Figure 2. Assessment form for the assessor to give feedback to the author

Variables and measures

Three variables were examined in this study: quality of peer-feedback provided, peer-feedback received, and question-generation. First, for the assessment of quality of peer-feedback provided, all the descriptive comments an assessor provided to the respective assessed items were analyzed against a pre-defined scheme with scores of 0 to 4 (the scheme is discussed in later section of this section) by a rater (having experience with both the focal science learning material and analysis procedure). The scores gained on each item in the same session were summed up (i.e., sessional score), and then divided by the number of implementation sessions (i.e., 4) as the average sessional quality of peer-feedback provided. For illustration purposes, as shown in Table 1, assessor A assessed a total of 10 questions during the four implementation sessions. The scores gained by each item in each session were summed up (i.e., 2+3+7+9=21), and then divided by 4 (i.e., 21/4= 5.25) as his or her average sessional quality of peer-feedback provided during the activity.

Table 1. An illustrative example of one assessor's scores on quality of peer-feedback provided

	Implementation #1		Implementation #2		Implementation #3		Implementation #4	
	Item	Score	Item	Score	Item	Score	Item	Score
Assessor A	1-1	1	2-1	2	3-1	2	4-1	4
	1-2	1	2-2	1	3-2	3	4-2	3
					3-3	2	4-3	2
Sessional score	2		3		7		9	
Average sessional score	5.25							

Second, for the quality of peer-feedback received assessment, all the descriptive feedback an author received from different assessors with regard to a specific question item was analyzed against the same scheme by the same rater, and averaged to yield feedback-received per item score (e.g., as shown in Table 2, (1+2)/2=1.5 for item 1-1, (2+0)/2=1 for item 2-1). The average scores gained on each item in the same session were then summed up (i.e., sessional score) across all sessions, and divided by the number of implementation sessions as the average sessional quality of peer-feedback received. For illustration purposes, as shown in Table 2, author A composed a total of 8 questions and received peer-feedback on 6 (i.e., no feedback received for items 1-2 and 2-2). For implementation session #1, author A received feedback from assessors B and C on item 1-1, but no feedback for item 1-2, and thus, the sessional score for implementation #1 is 1.5 (i.e., (1+2)/2+0=1.5). The average sessional score was then 2.5 ((1.5+1+3.5+4)/4= 2.5).

With reference to Nelson and Schunn's (2009) study on feedback and Yelon's (1996) open communication instructional principle with regard to providing feedback for students' work, the quality of online peer-feedback provided and received was evaluated using the same scheme in terms of the following five discrete levels: meaningless comments (0); general comments (1); specific comments, where strengths and weakness are identified

(2); identification and explanations for improvement (3); and explicit suggestions for further refinement of questions (4). For example, as shown in Table 3, for question #3, although specific comments (i.e., two strengths and one weakness) were provided, it did not offer any explanation for question-refinement, which were offered for questions #1 and #2. Also, for question #2, not only were areas for improvement identified and explained, but explicit suggestions were offered, although these were not for question #1. As such, the quality of peer-feedback provided for question #3 was scored "2," whereas that for question #2 was scored "4," and that for question #1 was scored "3," based on the 5-level scoring scheme. The higher the score is, the better quality the comment provided and received is. High quality comments tend to be more detailed, specific and instructive, as attested by Nelson and Schunn (2009), who indicated that feedback of this nature would lead to better performance.

Table 2. An illustrative example of scores on quality of peer-feedback received

	Implementation #1 ($2^a/1^b$)			Implementation #2 ($2^a/1^b$)			Implementation #3 ($2^a/2^b$)			Implementation #4 ($2^a/2^b$)		
	Item	Assessor	Score	Item	Assessor	Score	Item	Assessor	Score	Item	Assessor	Score
Author A	1-1	B	1	2-1	B	2	3-1	C	2	4-1	B	3
		C	2		D	0[c]		D	1		D	1
	1-2		x[d]	2-2		x[d]	3-2	E	2	4-2	C	2
Sessional score		1.5			1			3.5			4	
Average sessional score	2.5											

Note. [a] Number of questions generated by Author A; [b] Number of questions being assessed by peers; [c] The feedback was rated "0"; [d] No feedback received from peers.

Third, with regard to the quality of question-generation assessment, and in reference to Torrance's creativity index (i.e., fluency, flexibility, elaboration, originality) (1974), the cognitive levels proposed by King (1992), and the questions generated by students, a set of criteria was developed and operationally defined. Each question that the students generated during the question-generation activities was analyzed and scored against these criteria by the same experienced and qualified rater. Specifically, each question was graded along the following six dimensions:

- Fluency (0~3) assesses the correctness of wording and punctuation, clarity of meaning and logic, and relevancy of the constructed question.
- Flexibility (0~2) gauges the interconnectedness between the currently covered topic/unit and prior topics/units, and any self-derived examples.
- Elaboration (0~2) assesses the refinement of the questions in terms of creating scenarios for the question and using multimedia files to enhance understanding of the question.
- Originality (0~2) examines the uniqueness of a specific question as compared to those of peers in terms of innovative ways of formulating questions and content/ideas.
- Cognitive level (0~3) evaluates the cognitive levels demanded of the question-author, and these can be in terms of fact, comprehension or integration. Fact stresses the verbatim nature of questions from the learned materials, whereas comprehension indicates that students use their own words to define or describe the learned content. Integration shows that a link has been built across topics/units, and that explanations have been provided to build connections.
- Importance (0~1) evaluates the importance of the concepts assessed in the constructed question.

For example, as shown in Table 3, because all three example questions addressed the "correctness" and "relevance" elements, and the meaning of question #2 would be strengthened if the word "only" was included in the question (as suggested by one peer-assessor), it was scored "2" on the fluency dimension (due to not receiving a score on the "clarity" element), while questions #1 and #3 scored "3." Besides, as all three questions tested the currently covered topic/unit, but did not refer to concepts in prior topics/units, and only question #1 included self-derived examples, question #1 was scored "1" on the flexibility dimension, and questions #2 and #3 received a score of "0." Furthermore, as all questions #1~3 included self-created scenarios, each gained one point on the "scenario" element. Note that even though question #1 included a picture in one of the options, because it did not help to enhance the understanding of the item, as defined by the scoring scheme, no extra point was given for the "multimedia" element. Moreover, with more than one correct way to respond to the question (i.e., how to change the pitch of a guitar), the

combination approach used in question #3 was viewed as unique, as it was not seen in a question produced by any other student. As a result, among the three examples, question #3 is the only one that gained a score with regard to "innovative ways of formulating questions." Finally, as all questions touched on important concepts in the study material and were paraphrased, but none of them included any explanations with regard to connections across topics/units, all questions scored "1" on the importance dimension and "2" on the cognitive level dimension (i.e., comprehension). All scores gained on each of the dimensions per question per week were summed up and then divided by the number of implementation weeks (i.e., 4) as the average weekly quality of question-generation. The higher this score is, the better the overall quality of the student-generated questions.

To show the relation between the rater, the assessor, and the assessed, a figure was created with the inclusion of the criteria used by the respective parties/roles for evaluating the quality of question-generation and PA (see Figure 3). Finally, to establish inter-rater reliability, 10 pieces of student work (i.e., questions generated and comments received with regard to these) were randomly selected from each of the 8 participating classes from 2 out of the 4 implementation weeks. The samples (N =160) were evaluated by another independent rater, who was trained on the devised scheme and criteria before proceeding with the task. The results of the Pearson correlation between the two raters indicated that the coefficients were satisfactory (r =.94, p < .01 for PA and r =.73, p <.01 for question-generation).

Table 3. Examples of student question-generation and PA with reference to the respective scoring schemes

SQG examples	Quality of question-generation	PA examples	Quality of PA comments
Question #1: Spiderman sees people selling things at the night-market. He hopes to buy an item that conducts heat the fastest. Which of the following items should he buy?  (A) A steel spoon (B) Bamboo chopsticks (C) Paper (D) Plastic Answer key: (A)	• Fluency: correctness (1), clarity (1), the relevancy (1) • Flexibility: self-derived examples (1) • Elaboration: self-created scenario (1) • Originality (0) • Cognitive level: comprehension (2) • Importance (1)	+* The question is set in a context (night-market) with an interesting comic book character. +* "Underline" was used to highlight the main concept being tested in the question. +* All four options are items made of different materials with different levels of heat conduction. The question generated can assess students' level of understanding of the study content (i.e., heat conduction). -* Only the correct option has a picture in it. It may give the answer away.	Level 3: Identification and explanations for improvement
Question #2: John uses a heater in his room in winter. His mom says that the heater works on radiation. Answer key: False	Fluency: correctness (1), the relevancy (1) Flexibility (0) Elaboration: self-created scenarios (1) Originality (0) Cognitive level: comprehension (2) Importance (1)	+ A question with a meaningful scenario - A heater was the example included in the textbook. Other appliances frequently encountered in daily life (e.g., toasters) should be used to evaluate students' level of understanding (not memorization) on this topic. Based on the answer key provided by the author, because both radiation and heat convection play a role in how heaters work, the question should be revised	Level 4: Explicit suggestions for further refinement of questions

Question #3: Mary is practicing on her guitar, but finds the pitch is off. What can she do to change the pitch of her guitar? 1. Change the tightness of the strings 2. Change the sequence of the strings 3. Change the thickness of the strings 4. Change the color of the strings 5. Change the length of the strings (A) 123 (B) 125 (C) 135 (D) 124 Answer key: C	• Fluency: correctness (1), clarity (1), the relevancy (1) • Flexibility (0) • Elaboration: self-created scenarios (1) • Originality (1) • Cognitive level: comprehension (2) • Importance (1)	to be more concise: "…His mom says that how the heater works is ONLY based on radiation." + The presentation of the question is very special. Adds difficulty to the question. + All options are about the same length and based on a special aspect of the string, which makes it harder for students to get the correct answer simply by guessing. - All options have 1 as part of the answer.	Level 2: Specific comments where strengths and weakness are identified

Note. *+ = Identified as strengths by a peer-assessor; - = Identified as weaknesses by a peer-assessor.

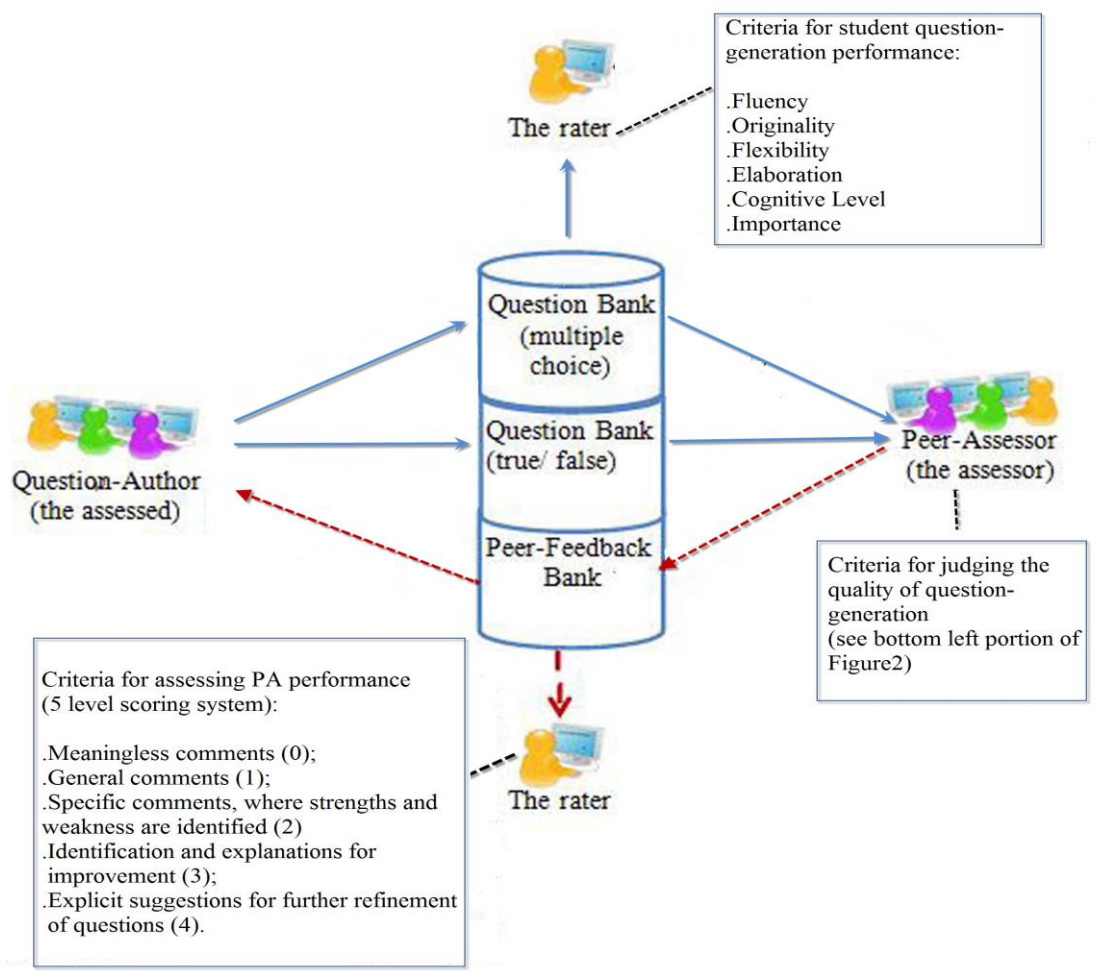

Figure 1. Relations between the rater, the assessor, and the assessed

Results

The correlations among variables, as well as the means and standard deviations of the examined variables, are listed in Table 4. As the assumptions with regard to multicollinearity and independent errors were not violated, as indicated by correlation between predictors (r = .13, p > .05), and all VIF (>1) and Durbin-Watson values (= 1.79), thus we can proceed to the hierarchical regression analyses.

Table 4. Correlations between examined variables and descriptive statistics (N = 213)

	Pre-assessment I[a]	Pre-assessment II[b]	Quality of peer-feedback received	Quality of peer-feedback provided	Quality of question-generation
Pre-assessment I [a]	1	.32**	.30**	.43**	.56**
Pre-assessment II [b]		1	.15*	.46**	.40*
Quality of peer-feedback received			1	.13	.37**
Quality of peer-feedback provided				1	.54**
Quality of question-generation					1
Mean (SD)	5.06 (2.98)	3.82 (3.11)	6.62 (3.75)	5.36 (3.02)	5.75 (2.27)
Maximum scores	14	19	19.38	20	13.25
Minimum scores [c]	0	0	0	0.75	1.00

Note. *p < .05; **p < .01. [a] The quality of question-generation at the training; [b] The quality of peer-feedback provided at the training; [c] indicates that the quality was rated to be of no quality based on the scoring scheme.

The predictive effect of the quality of provided and received online peer-feedback on the quality of question-generation

As shown in Table 4, the results from correlation analysis found significant relationships among the criterion variable, predictors, and students' pre-existing abilities (i.e., quality of students' question-generation and feedback provided during the training sessions). Thus, both pre-existing abilities were used as pre-assessments and entered into the regression model as step 1 to control for their effects before entering the predictor variables in the subsequent step(s).

The two-step hierarchical regression results showed that the quality of peer-feedback provided significantly predicted the quality of question-generation of the assessor (β =.32, p < .01), and thus students providing high quality feedback to their peers would be more likely to have high-quality questions.

Similarly, the quality of peer-feedback received, after controlling for the pre-assessment effect, had a significant predictive effect on the quality of question-generation of the assessed (β = .21, p < .01). In other words, this indicated that students receiving more specific and higher quality online feedback from the assessor would be more likely to have better quality questions.

The combined predictive effect of the quality of provided and received online peer-feedback on the quality of question-generation

A three-step hierarchical regression was conducted. As shown in Table 5, the quality of peer-feedback provided predicted a significant proportion of the variance with regard to the quality of question-generation of the assessor (R^2 = .44, F = 54.66, p < .01). In model 3, adding the variable of peer-feedback received significantly enhanced R^2 (R^2 change = .04, F = 16.53, p < .01), thus indicating that, together, the quality of online peer-feedback provided (β_{FP} = 0.33, p < .01) and received (β_{FR} = 0.21, p < .01) significantly predicted the quality of question-generation of the students who simultaneously played the role of both the assessor and assessed. In other words,

students' engagement in both roles was more likely to produce high-quality questions than when playing only one role.

Table 5. Hierarchical regression analyses of the quality of peer-feedback with regard to predicting the quality of question-generation

	Model 1			Model 2			Model 3		
Variable	B	SEB	β	B	SEB	β	B	SEB	B
Constant	3.21	.26		2.62	.27		2.02	.30	
Pre-assessment I	.37	.04	.48**	.29	.04	.38**	.25	.04	.32**
Pre-assessment II	.18	.04	.24**	.09	.04	.13	.08	.04	.11
FP				.24	.05	.32**	.25	.05	.33**
FR							.13	.03	.21**
R^2		.37			.44			.48	
F		60.76			54.66			48.18	
R^2 change					.07			.04	
F (R^2 change)					27.26**			16.53**	

Note. Dependent variable: question-generation; FP denotes peer-feedback provided to the assessed; FR denotes peer-feedback received from the assessor; $^{*}p < .05$, $^{**}p < .01$.

Discussion and conclusions

Several important findings were obtained from this work. First, the quality of online peer-feedback provided to the assessed was found to have predictive effects on the quality of the assessor's question-generation. In other words, an assessor who provided higher quality online feedback to his/her peers tended to produce better quality questions. Students who engaged in cognitive processes requiring observation of peer-produced artifacts, assessing their quality based on a set of criteria, and providing high quality feedback that assisted their peers to improve the quality of their work, were found to have greater improvements in their own work. This finding supports the results of studies such as Cho and Cho (2011), Li et al. (2010), Lu and Law (2012), and Lu and Zhang (2012), and further substantiates the learning benefits of providing peer-feedback.

Secondly, the quality of online feedback received from the assessor was also found to significantly predict the assessed students' quality of question-generation. When examined from the perspective of the assessed, students receiving higher quality online feedback appeared to then produce better questions. This result not only confirms what most researchers intuitively expect from PA (i.e., the positive effects of receiving peer comments) (Cho & Cho, 2011; Nelson & Schunn, 2009; Topping, 2010), although no supportive empirical evidence has been reported in the literature (e.g., Cho & Cho, 2011; Li et al., 2010; Lu & Law, 2012; Lu & Zhang, 2012), but also supports the findings of Nelson and Schunn (2009), suggesting that higher quality feedback (i.e., identification of problems along with clear suggestions for revision) can improve the quality of students' work.

Third, the combined predictive effects of online feedback provided to the assessed students and received from their peer-assessor on the quality of question-generation were confirmed in this study. As shown in the current study, students providing and receiving higher quality peer-feedback while serving as both the assessor and assessed was found to be related to a higher quality of question-generation as compared to students only engaging in one role. On the basis of this result, it may be inferred that the cognitive overload that the authors were initially concerned about was not present in this study. To elaborate, as implied by cognitive load theorists, if the tasks involved in playing both roles exceed the cognitive capacity of the learner, this will cause a state of cognitive overload, leading to diminished learning effects. Nevertheless, as shown in the results of the three-step hierarchical regression, student performance at question-generation did not decrease, but instead was significantly enhanced when compared to that seen when the students only provided feedback. Possible reasons for why playing both roles simultaneously did not result in diminished learning effects, as implied by the cognitive load theory, are offered.

As noted in the Introduction section, by playing the role of the assessor, students are encouraged to make objective judgments on the quality of the work they are viewing, and give constructive comments about it. On the other hand, when playing the role of the assessed, the comments provided by the assessor need to be studied, assessed, and integrated, if deemed appropriate. On the surface, the total number of tasks that needs to be attended to exceeds that involved in playing any one role (i.e., feedback-receiver or provider), and thus cognitive load theory suggests that the

students may be overwhelmed by simultaneously processing all of the information elements associated with the learning tasks (Paas et al., 2004). Nevertheless, if analyzed more closely it can be seen that both roles revolve around the set of criteria that are used for assessing the quality of question-generation.

To elaborate, efforts at evaluating the quality of an item and constructing descriptive comments as feedback would be most productive if they are centered on the set of criteria for the quality of question-generation. Similarly, processing and assessing the validity and usefulness of each piece of feedback that is provided would be easier if these criteria are also considered. Moreover, the provision of feedback requires the assessors to understand and apply the criteria for the quality of question-generation, and also helps the assessors to incorporate the information elements addressed in the criteria in a schema. With reference to the cognitive load theory that suggests students' schema availability influences the cognitive load imposed by a task (Paas et al., 2004), the schema of the criteria constructed during the feedback provision task may help to reduce the cognitive demand of the feedback reception task. Therefore, the tasks involved in being both a feedback-receiver and provider seem to be mutually beneficial, with students gaining fluency and mastery each time they engage in this practice. That is, engaging in the tasks associated with the role of a feedback-provider should help in the successful fulfillment of those associated with a feedback-receiver, and vice versa. As such, even though each of the tasks to be executed involves the activation and use of complex processes, as the essential concepts are more or less the same and help schema formation, it may be that its total cognitive demand did not exceed the cognitive capacity of the learners to the extent of negatively affecting student performance.

Finally, the significant results from the three-step hierarchical regression further indicates that primary school students are capable of accomplishing the tasks associated with being a feedback-receiver and provider simultaneously, and can benefit from doing this.

Scientific significance and implications of the study

This study contributes to the literature on PA and has the following empirical and methodological significance. First of all, related studies have mostly drawn their inferences based on participants' subjective perceptions of the learning improvements obtained as a result of feedback given by peers, as well as the participants' attitudes towards PA (Wen, Tsai & Chang, 2006). This study moves beyond such research by adopting objective measures to correlate the quality of PA with the quality of the work that is produced.

Secondly, by adopting the perspectives of both the assessor and assessed, and by differentiating both the effects of providing and receiving peer-feedback on the quality of produced work, their predictive effects, both individually and in combination, with regard to elementary students' quality of question-generation were empirically examined and supported.

Thirdly, because of the validation of these combined predictive effects, a long-held belief and practice associated with PA, based on the idea that learners benefit by playing the roles of both the assessor and assessed, was empirically substantiated for the first time.

Two suggestions are provided for instructional implementations. First, because the individual and combined effects of the quality of online peer-feedback provided and received were substantiated in the current study, ways to enhance the quality of peer-feedback should be explored in order to achieve the best results. Considering that feedback is important and is also a learned skill (Nelson & Schunn, 2009), possible ways to promote the provision and reception of high quality peer-feedback may include: context-specific assessment criteria regarding the target activity, training sessions with regard to the provision of quality feedback in accordance with a set of criteria, and whole-class feedback provided by the instructor highlighting exemplary PA practices, among others.

Second, instructors concerned about the possible cognitive overload that may result from students' playing multiple roles (i.e., the assessor and assessed) and completing multiple learning tasks (i.e., producing work, providing and receiving feedback) during PA can rest assured, as the findings of this study indicate that primary school students can benefit from engaging in the all the processes associated with the feedback-provider and receiver. As such, instructors should provide students with opportunities to both assess peer-generated work and receive peer-provided comments.

Limitations of this study

The current study examined the individual and combined predictive effects of online peer-feedback provided and received on fifth-grade students' quality of question-generation. As noted earlier in this work, with the support of networked technologies it is possible to easily carry out automatic, random assignment of students' work to be assessed and peer-feedback to be received. Another important point in the research design of this study is that fifth graders were recruited to participate for six consecutive weeks, and the assessors were given some freedom to choose with regard to which items to assess. As such, the generalizability of the results of this study to other contexts involving different age groups, or the employment of different media or work assignments, which may not be functionally equipped to support fluid interaction in PA, should be exercised with caution. Finally, although the students' pre-existing abilities in the focal tasks (i.e., question-generation and providing feedback) were controlled by including them in the regression models tested, the readers are advised to note that this study involved intact classes and differences among classes and between genders were not accounted for, and that the state of cognitive overload was inferred by statistical testing rather than empirical testing in this study.

References

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629–643.

Haladyna, T. M., & Downing, S. M. (1989). A Taxonomy of mulitple-choice item-writing rules. *Applied Measurement in Education, 2*(1), 37–50.

Hwang, G.-J., Hung, C.-M., & Chen, N.-S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology, Research and Development, 62*, 129-145.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126.

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

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

Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A Path model. *Journal of Educational Computing Research, 46*(3), 313–333.

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*, 257–275.

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

Nelson, M. M., & Schunn, C. D. (2009). The Nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401.

Osterlind, S. J. (1998). *Constructing test items: Mulitple-choice, constructed-response, performance and other formats.* Boston, MA: Kluwer Academic Publishers.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.

Piaget, J. (1926). *The Language and thought of the child*. New York, NY: Harcourt Brace & Co.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A Review of the intervention studies. *Review of Educational Research, 66* (2), 181–221.

Taiwan Elementary and Educator Community. (2014). *The Core competency.* Retrieved from http://teach.eje.edu.tw/9CC2/9cc_basic.php

Torrance, E. P. (1974). *Torrance tests of creative thinking*. Bensenville, IL: Scholastic Testing Service.

Topping, K. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*, 339–343.

Topping, K., & Ehly, S. W. (Eds.) (1998). *Peer assisted learning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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.

Wen, M. L., Tsai, C. C., & Chang, C. Y. (2006). Attitudes toward peer assessment: A Comparison of the perspectives of pre-service and in-service teachers. *Innovations in Education and Teaching International, 43*, 83–92.

Yelon, S. L. (1996). *Powerful principles of instruction*. White Plains, NY: Longman.

Yu, F. Y. (2012, November). Learner-centered pedagogy + adaptable and scaffolded learning space design—Online student question-generation. In *International conference on computers in education* (pp. 26-30). Singapore.

Yu, F. Y. (2009). Scaffolding student-generated questions: Design and development of a customizable online learning system. *Computers in Human Behavior, 25*(5), 1129–1138.

Yu, F. Y., & Liu, Y. H. (2005). Potential values of incorporating multiple-choice question-construction for physics experimentation instruction. *International Journal of Science Education, 27*(11), 1319–1335.

Yu, F.Y., Liu, Y. H., & Chan, T. W. (2005). A Web-based learning system for question-posing and peer assessment. *Innovations in Education and Teaching International, 42*(4), 337–348.

Yu, F., & Wu, C. (2013). Predictive effects of online peer feedback types on performance quality. *Educational Technology & Society, 16*(1), 332–341.