Toward Enhancing the Validity in Multi-Peer Assessment: An Approach Based on Students' Speed-Response-Question Performance

Yanqing Wang[1,*], Yong Wang[2,1], Haocheng Yuan[1], Eddy N Mandimby[1], Shiying Hu[1]

[1]School of Management, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

[2]Heilongjiang Institute of Teacher Development, Harbin, Heilongjiang 150080, China

[*]Corresponding author. Email: yanqing@hit.edu.cn

ABSTRACT In many peer assessment systems, practitioners found that the scorings of students' work by peers are much higher than their work deserves. In order to improve the validity of student's scoring during the multi-peer assessment process, this paper introduces practical solutions towards the above process, it also applies the relevant solutions into the peer assessment process, it passingly finds the qualified "little teaching assistants" along the grading process, and it uses these "little teaching assistants" for a higher efficiency towards grading students as well as enhancing the validity of the student's scoring results.

1. INTRODUCTION

The teaching response system or the audience response system is an interactive teaching approach based on the utilization of wireless communication technology. A combination of interactive teaching techniques, interactive learning activities, interactive teaching evaluations will be implemented in the classrooms in order to effectively enhance a higher participation level and a higher enthusiasm level from learners, which is considered to be a new research field of educational technology. Since the early 1960s, the United States and the United Kingdom, have attached very great importance to the research and the development of the teaching response system, which is mainly used for educational, military and business purposes. With the popularization of higher education, the teaching response system can effectively stimulate better classroom management especially for the large classes, which can effectively stimulate a better learning environment, a better interaction between teachers and students, and a moderate promotion to the pedagogical performance [1]. Researches in teaching response system are moderately increasing which is playing a very significant role into the advancement of the field, its teaching methodology as well as the offer of innovative learning tools to promote a greater teaching reform that can effectively strengthen more in-class questioning, more feedbacks and more participation from learners. The optimization of the learners' attention and focus has a positive significant relationship with the change of teaching methods used, the evaluation process used, peer teaching incentives and the implementation of the problem-based teaching methodology.

Generally speaking, there are two means of in-class speed-response-question (SRQ), including *a) Raising-hand SRQ*, which is the traditional audience engagement method through raising your hands to answer questions [2] and *b) Online SRQ*, which is performed through online software. The four years practice of hands-up to answer questions, although it provided "an effective teaching interaction and high visibility in term of the level competitiveness", it has been highly noticed that "the feeling of being shy, the fear of being a subject of mockery, the fear of making mistakes, the fear of losing face" and other weaknesses are growing among Asian students in the main reason that respect elders are judged as being a fundamental value, therefore, the teachers have the highest authority that is not meant to be challenged [3-4]. As a result, it decreases the contribution of in-class SRQ activities. Therefore, from this semester on, the implementation of the "in-class online

question" will be a major leap into a greater advancement of the field of educational technology. In fact, peer assessment is known for tremendously helping faculties in terms of workload [5] as well as the increase of learning outcomes [6].

In order to have better results in the implementation of the online classes, supporting hardware and/or software that comes at a premium cost, a premium configuration and maintenance costs required are the main reasons why it is not currently widely used yet. Moreover, the smartphone usage rate among Chinese college students has already reached 100%. Therefore, the understanding of how to effectively use mobile phones for greater interaction, the shift from the traditional learning pattern in classrooms to the virtual learning platforms which is the Campus Mobile Classroom that created more needs for education [7]. The main pedagogical interactive method is simply classroom questioning. The digitalization, networking, and mobilization are highly required in information technology implementation [8]. Thus, software selection becomes the last task to do. Two popular software tools are optional. *1) app.* To be customized, professionally developed with a user-friendly interface. However, its development and maintenance, as well as upgrading costs, will be at a premium level. In particular, the mainstream mobile operating system which includes *Android* and *iOS*, the students' usage ratio of those operating systems will be approximately 3:1. Therefore, professional development of apps remains secondary not an ideal choice; *2) WeChat official accounts.* This model is currently used for educational researches: teachers create a WeChat official account from Tencent Company, and students use the WeChat platform to answer SRQ.

2. About Online Peer Assesement

Topping points out that peer review, also known as peer assessment or peer feedback, is the evaluation of certain academic works or achievements from peers within roughly the same level of academic background where the evaluation is based on the quantity, level, value, function, and quality of peer works. As a formative assessment, peer revision is not only limited to the evaluation of certain assignments but could be a learning opportunity as well as. The modern educational-related challenges are highly focused on designing self-directed and collaborative learning activities [9], therefore, the peer assessment implementation responds positively on getting the learners' attention [10] compared with evaluations established by teachers, peer assessments enable students to fully engage themselves within the overall learning process. However, certain reluctance from learners to participate in certain peer assessment activities will be highly noticed where certain students are skeptical about the credibility of these evaluations mainly due to their peers' doubtful academic abilities, the quality of the relationship among peers as well as the level of accountability. Nonetheless, once the above doubts are overcome and the evaluation is accurately implemented, peer reviews are mutually beneficial to both teachers and learners. Furthermore, studies have consistently found out that peer revision is highly beneficial for both the assessors and the assesses [11-13] and the quality of the reviews will be highly dependent on the student's attitudes. In fact, certain studies pointed out that practitioners having a positive attitude towards peer feedback are willing to provide revision exceeding the expected standards [10]. The engagement level will be increasing and the carelessness level decreasing when practitioners hold a positive attitude towards it [14]. Interactive peer assessment, while reducing the number of burdens in favor of teachers, helps learners to understand the learning process of other peers as well as the identification of their own strengths or weaknesses. Moreover, students' independent thinking ability, as well as their independent learning, reflection and questioning, critical thinking, and other higher-level thinking skills, could be enhanced throughout the peer revision process.

With the advancement of online learning, student evaluation has become one of the most arduous tasks for teachers. In response to the thousands of assignments to be reviewed, the introduction of peer assessment mechanism has become a necessary measure for a greater learning environment. During the face-to-face peer reviews, students tend to be more focused on the social dynamics than providing accurate feedbacks, such as whether or not they are perceived as mean or as an embarrassment to others, rather than supplying adequate feedbacks [15-16]. In fact, the learning

behaviors are highly connected to social-based psychological safety where people tend to align themselves with negativity from the team members when reviewing peers [17]. Moreover, prior studies suggested that social pressure or negative attitude could discourage the usage of peer assessment [18-19]. However, other studies suggested that practitioners viewed peer assessment as an effective tool and they would recommend its usage to others [20] when rightfully implemented.

Therefore, online peer-reviews have been currently implemented so that authors and reviewers are anonymous. Lewin carried out empirical research on peer assessment where the results of peer assessment were positively correlated with the results of teacher evaluation. During the implementation of the direct peer assessment, the mitigation of the evaluator's emotional factors and personal interests can significantly improve the quality of peer assessments. In general, based on the scientific evaluation criteria, the online peer assessment is carried out by the combination of scoring and emergency response, with a better reliability and validity level. In the meantime, it can promote learning outcomes. Wang et al proposed an incentive model of peer assessment. Their empirical study shows that the incentive model increases the chance for automatic anomaly detection and manual correction of peer assessment, improves the evaluation fairness and reliability of students and, in order to improve the validity of the evaluation, further exploration is necessary [21].

3. Experiment

3.1. Research Background

This study considers the teaching objectives, teaching content, and the analysis of learners' characteristics and it considers the design and implementation of peer assessment system learning activities. In order to explore the impact of peer assessment on evaluation results, this study adopts the quasi-experimental method.

This study mainly focuses on Big Data majors, from the second to third-year undergraduate programs of a Chinese university. Therefore, two classes were selected from the first semester of the 2019-2020 academic year in which both classes were targeted to have the same curriculum, teaching objectives as well as the same instructor. Moreover, learners are required to meet certain criteria such as holding at least one year of the online learning experience, being able to master the online learning platform, and have high information literacy. Prior to officially attending the courses, students were assigned to view certain information on the online learning platform, to complete the tests and assignments independently with its submission before the actual deadline. However, these participants did not have prior peer assessment experience.

3.2. Relevant Definitions

The sole purpose of this experiment is the enhancement of the differences between the peer assessment with and without SRQ. A study based on an online peer assessment system involved 81 students in two classes.

(1) Question: Questions were precisely designed by the teachers beforehand with a difficulty coefficient ranging from medium to high while the students' comprehensive ability is thoroughly assessed within 25 minutes. There are four types of questions that have been used for the study: open question, multiple-choice question, output giving question and error finding questions, as shown in Table 1.

Table 1 Samples of a question in SRQ repository

Sample #1	Sample #2
Type: *open-ended*	Type: *choice*
Content: Please explain the difference of a float variable from the perspective of a screen display, storage, and value in a program.	Content: In which of the following situations, could data type get a promotion? *A. variable with higher data type is assigned to another with lower data type* *B. mixed calculation* *C. manual conversion with cast*
mark: *2*	mark: *2*
Sample #3	**Sample #4**
Type: *output giving*	Type: *error finding*

Content:
#include <stdio.h>
void main()
{
   int counter = 0;
   counter++;
   printf("%d\n", counter--);
}
	Content:
	#include <stdio.h>
	void main()
	{
	   int var5;
	   var5 = 1;
	   var5++2;
	   printf("");
	}

mark: *1*	mark: *1*

(2) Question bank (repository): Preparation of emergency questions based on the chapters studied as well as the whole course and the question library.

(3) Testing (quiz). It is highly dependent on the difficulty coefficient and the class schedule, 10 to 16 questions were precisely selected from the question bank that students are required to answer them within around 25 minutes.

(4) Answer score. According to the order of correct answers, the top three students with the correct answers will get corresponding scores.

(5) The score of the successful respondent. It is a cumulative score, i.e. the score of the *n*th time is the summary of the previous *n*th times.

(6) Student expert. They are successful students, who had excellent performance in SRQ activity. They got "experts" calling and assigned much higher scoring weight than other students. Generally, we choose the top one third as student experts.

3.3. Process

The process includes four steps as follows.

(1) Getting ready. Students are assigned to first get familiar with the platform within the WeChat official account, get access to the public platform emergency response interface where certain interactions with the teachers are expected. Then the questions will be displayed on the screen where students are assigned to answer certain questions on the WeChat public platform (see Figure 1).
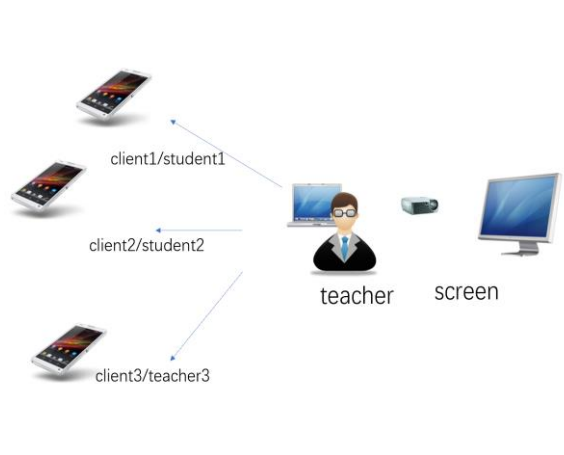
Figure 1 SRQ process

(2) SRQ testing (quiz). After the teacher starts SRQ, students can provide their answers, submit and store the teacher's questions in the WeChat public platform database. The platform database presents the answers and sorts them according to the time. The teacher verifies the results publicly. The top three students who give the correct answers are confirmed and checked/marked by the teacher. That means these three students obtain the corresponding scores right away. The above action is repeated until all prepared questions are completed by the teacher and students (see Figure 2).

(3) Ranking students by accumulative scoring. Successful respondents, referred to as the student experts, will play a major role within the adjustment process (see Figure 3).

(4) Model of weighted average. As 4-by-4 assignments are made, depending on the participant's task completion level (mark) which could be broken down into four effective sub-models accordingly, see Table 2.
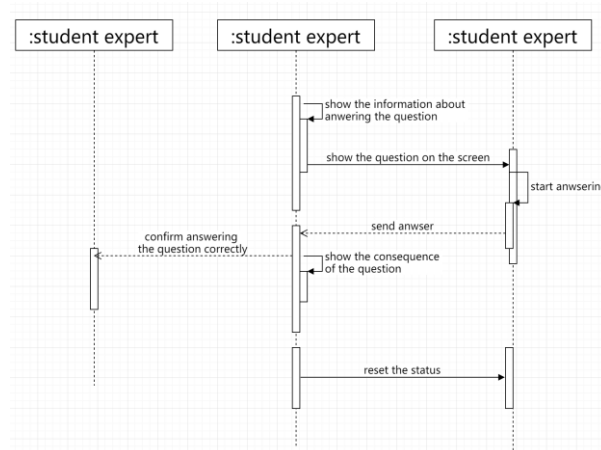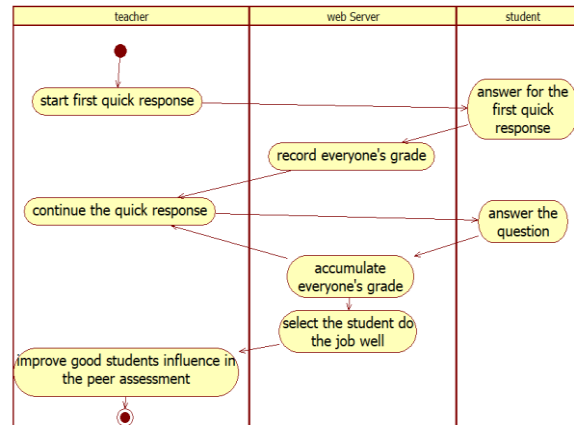

Figure 2 SRQ flowchart

Figure 3 SRQ and peer assessment flowchart

Table 2 Adjustment weights by number of scorings and number of experts

# of scorings	# of experts	weight 1	weight 2	weight 3	weight 4
4	1	52	16	16	16
	2	38	38	12	12
	3	30	30	30	10
	4	25	25	25	25
3	1	60	20	20	
	2	43	43	14	
	3	34	33	33	
2	1	75	25		
	2	50	50		
1	1	100			

Note: in each row, the higher weight is for an expert student and the lower weight for a common one.

For example, in the 4-by-4 assignment, as to the scorings for one students' work, there are four different possibilities, i.e. there are from 1 through 4 experts among the four reviewers (see the upper four rows in Table 2). Thus, the calculation of the overall task scores will be on the weighted average basis instead of the simple average one. Therefore, let the student experts evaluate other people's scores from the weighted average basis. Other students will be assigned to thoroughly follow similar procedures in order to score the student experts.

In the case that the importance of "little teaching assistants" is increased over time, and the validity of assessment respectively increases, so as to realize the process of weighting quantification.

4. Analysis

From in-class observation, interview, and data analysis, we found this approach is practical and effective.

(1) SRQ provides a much greater emotional engagement. In this study, the overall students' learning process was observed when attending the courses. Consequently, a much higher positive emotional pattern was experienced by learners who were engaged in the usage of preemptive responses than learners who were engaged in the usage of permissive ones. However, in a broader sense, the overall online peer review was enjoyable.

(2) The overall score adjustments are majorly negative. That is effectively restraining the phenomenon of "scorings to students' work by peers are much higher than their work deserves" (see Figure 4). In this study, two courses are involved. From Figure 4, it is found that the majority of tasks (assignments or projects) got more decrement than increment. Similar to many other peer assessment systems, students gave much more scores to peers' work than their work deserved. Thus, an obvious score decrement has validated our approach in this study is effective.
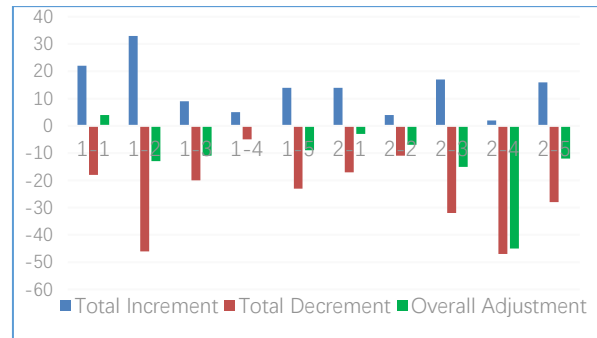
Note: "1-2" in x-axis means task 2 in course 1, and so on

Figure 4 Overall score adjustments to students' tasks

(3) The unexpected training of little teaching assistants. Surprisingly, the implementation of the online peer assessment from SRQ activity allowed a greater number of "little teaching assistants" to improve certain areas of the study, such as improvement on the assessment results, and the assessment results from the SRQ basis were far greater than the control group (identical teacher's another class without SRQ activity) without SRQ setting.

5. Findings AND Discussion

The study was mainly designed to provide a greater understanding of how peer assessment works and to provide a greater amount of accuracy in terms of students' grades where different grading systems were involved. In fact, the maintenance of a higher quality peer assessment, as well as a higher reliability level has constantly been the focused point of researchers and practitioners. However, it has been argued that a reliable peer assessment implementation is heavily dependent on the manner that it was incorporated into the learning process [22].

Since the recent technological revolution, our life dramatically has changed on a daily basis, especially in certain ways of doing things. Knowledge acquisition was no stranger to such a revolution. Therefore, technology could be a major tool to help students acquire a greater amount of knowledge and share experience, compared to the traditional way of simply attending classes. At the initial stage of this study, our research group has been relying on several computer software concepts such as *C Programming* and *Object-Oriented Programming*, committed to the design and the evaluation of online education. However, along the peer feedback process, the majority of learners had a low interaction level and a low behavioral input leading them to casually comment "meaningless comments", such behavior is assumed to be caused by either having a low level of knowledge or a high level of carelessness. Hence, the introduction of the online peer assessment utilizing SRQ that developed greater teaching practices. Moreover, this research takes into account the core course *C Programming* as an example, designs online peer assessment learning activities with quick response, and uses a quasi-experimental research method to explore its impact on learners' online recognition and evaluation results.

Our study promoted a newer approach to the pedagogical performance of knowledge acquirers. In fact, the intersection of learning and technology, in itself, is a very innovative approach to learning and the insufficiency in terms of researches in the field is highly noticed. Our study is limited to the understanding of university knowledge acquirers where further researches on another type of knowledge acquirers remain necessary. Furthermore, the online peer assessment from the SRQ basis helps to improve the validity of the evaluations. From the results of the level test after the implementation of the teaching, the level of the experimental group is significantly higher than that of the control group.

In addition, learners are more in favor of an online peer assessment, and the evaluators scoring high grades "help to improve the validity of evaluation" as well as the improvement of the learning performance. Combined with the reflection of learners and the teaching reflection of researchers themselves, it is highly probable that WeChat can't integrate with the existing teaching information

system. Needless to say that teachers logged into the database with a high-frequency level throughout the whole study, in the main reason that the web technology is more mature than *app* technology for many years, and the cost of developing web application is far lower than the cost of developing app, and the compatibility of web application (using browser) is very high. In the near future, we will choose this kind of technical means of web-based in-class SRQ system, which will improve the learning outcomes greatly in the following teaching practice and research hopefully.

Acknowledgment

References

[1] H. Li, J. Zhang, Current situation and hot spot of application research on teaching response system, China Educational Technology, 5 (2012) 128–134.

[2] Y. Wang, Redesigning flipped classrooms: A learning model and its effects on student perceptions, Higher Education, 78(4) (2019) 711–728.

[3] X. Xie, Why are students quiet? Looking at the Chinese context and beyond, ELT Journal, 64(1) (2009) 10–20.

[4] J. P, Willingness to communicative in the Chinese EFL classroom: a cultural perspective, in: L. J (Ed.), English language teaching in China: New approaches, perspectives and standards, London, New York: Continuum, 2007.

[5] R. F. Rubin, T. Turner, Student performance on and attitudes toward peer assessments on advanced pharmacy practice experience assignments, Currents in Pharmacy Teaching and Learning, 4(2) (2012) 113–121.

[6] C. Murakami, Turning apathy into activeness in oral communication classes: Regular self- and peer-assessment in a TBLT programme, System, 40(3) (2012) 407–420.

[7] X. Zhang, Design and application of the APP of the teaching answer catcher, International IT Media Brand, 39(4) (2018).

[8] S. Wang, et al, Review of educational APP research, Journal of education institute of Jilin province, 5 (2014) 103–104.

[9] J. P. Voogt, et al, Technological Pedagogical Content Knowledge Review of the Literature, Journal of Computer Assisted Learning, 29(2) (2013) 109–121.

[10] Y. Zou, et al, Student Attitudes That Predict Participation in Peer Assessment, Assessment and Evaluation in Higher Edu. 43(5) (2018) 800–811.

[11] E. C. Berg, The Effects of Trained Peer Response on ESL Students' Revision Types and Writing Quality." Journal of Second Language Writing, 8(3) (1999) 215–241.

[12] T. M. Paulus, The Effect of Peer and Teacher Feedback on Student Writing, Journal of Second Language Writing, 8(3) (1999) 265–289.

[13] M. M. Patchan, et al, Writing in Natural Sciences: Understanding the Effects of Different Types of Reviewers on the Writing, Journal of Writing Research, 2(3) (2011) 365–393.

[14] M. M. Patchan, et al, The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions, Journal of Educational Psychology, 108(8) (2016) 1098–1120.

[15] M. Christianakis, "I don't need your help!" Peer status, race, and gender during peer writing interactions, Journal of Literacy Research, 42(4) (2010) 418–458.

[16] S. Peterson, Peer response and students' revisions of their narrative writing, L1-Educational Studies in Language and Literature, 3(3) (2003) 239–272.

[17] A. C. Edmondson, Psychological Safety and Learning Behavior in Work Teams, Administrative Science Quarterly, 44 (1999) 350–383.

[18] D. K. Cohen, Educational Technology and School Organization, in: Technology in Education: Looking toward 2020, 1988, pp.231–264.

[19] L. M. McNeil, Contradictions of Control: School Structure and School Knowledge, New York: Routledge, 1988.

[20] P.Vickerman, Student perspectives on formative peer assessment: An attempt to deepen learning, Assessment & Evaluation in Higher Education, 34(2) (2009) 221–230.

[21] Y. Wang, et al, Towards motivating participants to assess peers' work more fairly: Taking programming language learning as an example, Journal of Edu. Comp. Res. 52(2) (2015) 180–198.

[22] L. Schuwirth, Optimising new modes of assessment: in search of qualities and standards, Tijdschrift voor Medisch Onderwijs, 5(23) (2004) 250–251.