# Assessing and Supporting Argumentation with Online Rubrics

Jingyan Lu[1] & Zhidong Zhang[2]

[1] Faculty of Education, The University of Hong Kong, Hong Kong, China

[2] College of Education, The University of Texas at Brownsville, Texas 78520, U.S.A

Correspondence: Jingyan Lu, The University of Hong Kong, Pokfulam Road, Hong Kong, China. Tel: 852-2241-5450. E-mail: jingyan@hku.hk

## Abstract

Writing and assessing arguments are important skills and there is evidence that using rubrics to assess the arguments of others can help students write better arguments. Thus, this study investigated whether students were able to write better arguments after using rubrics to assess the written arguments by peers. Students in 4 secondary 4 classes at a publicly funded Hong Kong high school used an online assessment system to assess the arguments of peers for one year. Students first used a rubric to assess arguments along four dimensions: claims, evidence, reasoning, and application of knowledge. Then they compared their assessments with assessments by their teachers using the same rubrics. Data included student-teacher agreements on rubric dimensions, students' evaluation comments, and their perceptions of the assessment activity. Results indicated that the quality of students' written arguments could be predicted based on the number of student-teacher agreements on the rubrics dimension of evidence and on the number of students comments identifying problems and reflecting on assessment. This study shows that providing students with rubrics for assessing the written arguments of peers can lead them to write better arguments.

**Keywords:** rubric based assessment, assessing argument, argumentation model, online assessment, peer assessment, peer feedback

## 1. Introduction

Being able to write and to assess arguments competently is important in school for constructing and evaluating knowledge and in daily life for exercising the rights and duties of responsible citizenship. Although, it is recognized that the skills involved in effective argumentation ought to be taught in school (Driver, Newton, & Osborne, 2000; Nussbaum, 2002) they rarely are at least in any systematic way. This is in part because teachers have seldom been taught to do so. It is therefore not a surprise that many students (Knudson, 1992), including many recent high school graduates are unable to competently produce and assess arguments (National Assessment of Educational Progress, 1998; National Science Board, 2006).

Educational researchers have traditionally focused on arguments in math and the natural sciences (Aberdein, 2005; Erduran, 2007), more recently they have begun focusing on arguments in the social sciences and the humanities (Larson, Britt, & Kurby, 2009). Although, many studies have examined how students construct arguments (Chang & Chiu, 2008; Driver, et al., 2000; Li & Lim, 2008; Wu & Tsai, 2007), few have looked at how they assess them (Hagler & Brem, 2008; Kuhn, 2005; Larson, et al., 2009; Lu & Lajoie, 2008; Sadler, 2004) despite evidence that doing so enhances learning by involving students more deeply in the learning process (Gielen, Peeters, Dochy, Onghena, & Struyven, 2009; Goldstein, Crowell, & Kuhn, 2009). Further, few studies have focused on the use of rubrics in assessing arguments.

Students are now able to assess the work of peers online which, unlike face-to-face assessment, allows them greater freedom to review their own feedback and to compare it with teacher feedback. Further, online systems allow teachers to construct rubrics for students use in assessing arguments. Thus, the students in this study used an online assessment system and teacher generated rubrics to evaluate the written arguments of peers and then to compare their assessments with those of their teachers. They also had opportunities both to reflect on their assessment experiences and to compare their own written arguments with those of their peers. The study sought to determine whether students' online peer assessment activities and reflections lead them to write better arguments.

## 2. Literature Review

Research has demonstrated that the development of argumentation skills can promote productive thinking (Nussbaum, Winsor, Aqui, & Poliquin, 2007), reasoning (Hahn & Oaksford, 2007; Lajoie, Greer, Munsie, Wilkie, Guerrera, & Aleong, 1995), problem solving (Chiu, 2008), decision making (Karacapilidis & Papadias, 2001; Lu & Lajoie, 2008) and knowledge construction (Jamaludin, Chee, & Ho, 2009).

Although, we now know something about how students develop the ability to construct and assess arguments we have also discovered that developing the ability to do so is no easy task (Knudson, 1991). It has been found that graduating high school students are often unable to competently produce or assess arguments (National Assessment of Educational Progress, 1998; National Science Board, 2006). A number of reasons have been suggested for why this is the case. Some researchers have suggested that it may be because our pedagogies are inadequate (Jonassen & Kim, 2010). Bereiter and Scardamalia (1982) have suggested that students may lack appropriate schemata while Kuhn (1991) and Perkins, Farady, and Bushey (1991) have suggested that students may not get enough formal instruction or practice. In response to Bereiter and Scadamalian's (1982) suggestion the students participating in this study were given rubrics which they then used to assess the written arguments of their peers. The students then compared how they assessed the arguments of their peers with how their teachers assessed them. In this way, students were introduced to the appropriate schemata through the rubrics they used to assess the arguments of their peers.

The remainder of this literature review will focus on argumentation skills and how online peer assessment can help foster their development.

### 2.1 Models of Argumentation

The most influential model of argumentation in educational research was developed by Toulmin (1958) who identified six parts of arguments: (1) claims, (2) evidence, (3) warrants, (4) backing, (5) qualifiers, and (6) rebuttals. In this study we found the teacher created a rubric that incorporated the first three elements of Toulmin's model of argumentation which are claims, evidence, and warrant.

According to Toulmin (1958) an argument involves the movement from evident to a claim through a warrant. A claim is the conclusion of an argument. It is an assertion about some issue or phenomena that the arguer wants others to accept. Claims may vary in complexity from simple popularly held beliefs to complex scientific theories. The ability to make clear claims is developmental (Knudson, 1992). Evidence consists of facts or examples introduced to support a claim (Kuhn, 1991; Toulmin, 1958). Students often fail to provide sufficient evidence for their claims (Kuhn, 2001; Walton, 1996). Pragmatically, the availability and the strength of evidence can determine how well students justify their arguments (Brem & Rips, 2000; Kuhn, 2001). Warrants are general statements serving to link the evidence to the claims they support. They are seen as the ability to reason. Students often construct arguments in which they fail to explicitly link evidence to the claims they support (Knudson, 1992; Kuhn, 1991). Students, who are unable to distinguish between arguments in which evidence and claims are properly linked from those in which they are not, are unable to construct good arguments and to effectively assess the arguments of others (Larson, et al., 2009). Finally, not listed in the Toulmin's model but very important to argumentation is the knowledge. Constructing and assessing good arguments involves understanding, elaborating, and discussing key concepts and knowledge. Equipping students with key concepts and knowledge is also seen as a basis of effective argumentation.

The development of effective argumentation skills is both a means and a goal of education. Research indicates that students acquire new perspectives and understandings by constructing arguments in different subject domains. For instance, in the natural sciences, constructing and assessing arguments can enhance students' conceptual and epistemic understanding and can render scientific reasoning visible (Chi, Slotta, & de Leeuw, 1994; Duschl & Osborne, 2002). Similar result have been found in the social sciences and the humanities (Wiley & Voss, 1999). Writing arguments leads to better conceptual understanding than writing narratives, summaries, or explanations (Wiley & Voss, 1999). Learning activities that involve solitary or collaborative argumentation can lead to better knowledge gains than learning activities that do not (Asterhan & Schwarz, 2007).

### 2.2 Assessing Arguments

Students exercise the same argumentation skills in assessing arguments as they do in constructing them and although, school is where they should develop these skills little research has focused on how this happens. One promising approach involves the use of rubrics to evaluate and enhance student learning (Andrade, 2000; Jonsson & Svingby, 2007). Students who use rubrics to assess arguments have more consistent and reliable argumentation skills (Jonsson & Svingby, 2007) and construct better arguments. Larson, Britt, and Kurby (2009)

found that students who used rubrics to evaluate arguments and who received immediate feedback, improved their ability to judge the quality of arguments.

Scoring rubrics are rubrics employ descriptive scales and using such rubrics can provide students with a clearer understanding of what is important and can help them to evaluate the strengths and weaknesses of their work (Andrade, 2000; Moskal, 2000). Rubric-based peer assessment can scaffold argumentation skills by providing students with scales for assessing the features of arguments (Kuhn & Udell, 2003; Royer, Cisero, & Carlo, 1993). Thus, using rubrics to assess the arguments of peers can lead students to reflect on their own arguments and apply the same rubrics in writing them.

A number of studies have provided evidence that rubric-based peer assessment enhances student learning. First-year psychology students reported that using rubrics to grade the work of peers motivated them to think and learn more effectively (Falchikov, 1986). Similarly, first-year undergraduates reported that peer assessment enhanced their critical thinking, sense of structure, and learning (Orsmond, Merry, & Reiling, 1996). Stefani (1994) reported that students, who participated in developing a marking rubric for lab assignments became more reflective and successful learners. Hughes (1995) also reported that first-year undergraduates improved their performance by using detailed marking schedules for peer-marking. Research indicates that self-checking based on evaluative feedback as opposed to solitary practice enabled students to assess arguments more effectively (Larson, et al., 2009).

*2.3 Types of Feedback*

Students can use rubrics both to grade and to provide feedback on the work of peers. The feedback can involve commenting on the work which can involve reflective engagement (Falchikov & Blythman, 2001). Peer feedback has been found to improve the learning of both the assessor and the assessee (Li, Liu, & Steckelberg, 2010; Topping & Ehly, 2001; Xiao & Lucking, 2008). For instance, it can sharpen the critical thinking skills of assessors and it can provide timely feedback to assessees. This study focused on the effects of feedback on assessors as opposed to assessees. Assessors may summarize arguments, identify problems, offer solutions, and explicate comments. In so doing assessors may increase the time they spend thinking about, comparing, contrasting and talking about learning tasks (Topping, 1998). Further, assessors may review, summarize, clarify, diagnose misconceived knowledge, identify missing knowledge, and consider deviations from the ideal (Van Lehn, Chi, Baggett, & Murray, 1995). Assessors who provide high quality feedback have been found to have better learning outcomes (Li, et al., 2010; Liu, Lin, Chiu, & Yuan, 2001). For instance, Tsai, Lin and Yuan (2001) found that pre-service teachers who provided more detailed and constructive comments on the work of their peers performed better than those who provided comments that were less detailed and constructive. Topping, Smith, Swanson and Elliot (2000) found that assessors not only improved the quality of their own work but also developed additional transferable skills.

Peer feedback was carried out online and the next section reviews the literature on online assessment systems.

*2.4 Online Assessment*

Online assessment systems have changed the assessment process (Tsai, 2009; Tseng & Tsai, 2007) by enabling students to submit and store assignments, communicate with peers and teachers, and review and reflect on feedback. For example, "NetPeas" allows students to upload and modify assignments, assess the work of peers and file complaints (Lin, Liu, & Yuan, 2001). "Group Support System" allows students to discuss assessment criteria and carry out collaborative assessments (Kwok & Ma, 1999). Online assessment systems also affect learning. Tseng and Tsai (2007) found that 10th graders improved the quality of their projects by exchanging online feedback with peers. Yang (2010) designed an online peer review system that allowed students to observe and learn from each other by modeling, coaching, scaffolding, and reflecting during the writing process. She also found that the system helped students communicate with peers, review and assess their written assignments, and reflect on and revise their own work.

## 3. Research Questions

This study investigated whether using an online assessment system to assess the written arguments of peers can lead students to write better arguments. The online system allowed students to compare how they assessed the arguments of peers with how their teachers assessed the same arguments. We hypothesized that comparing their assessments of the written arguments of peers with those of their teachers would lead students to reflect more deeply on criteria for evaluating the parts of assignments and then to reflect more deeply on their own written arguments. We also wanted to explore how assessing and reflecting on different parts of arguments affected learning outcomes. Do student-teacher agreements on assessments of peer arguments affect the learning

outcomes of assessors? Which types of assessment comments are most effective in learning? These issues led to three research questions.

1)    How do student-teacher agreements with respect to the assessment of the arguments of    peers influence the learning outcomes of assessors?

2)    How do the number and types of assessment comments influence the learning outcomes of assessors?

3)    How does reflecting on the online assessment of arguments influence the learning of    assessors?

## 4. Method

### 4.1 Participants

One hundred and twenty-one 13-14 year old secondary four students in a publicly-funded Hong Kong high school participated in the study. The students were from four different classes of approximately 30 students each. There were 43 girls and 78 boys. The school was chosen as a convenience sample from schools participating in a university-school partnership project involving the use of online platforms in teaching and assessing Liberal Studies, a core course in Hong Kong's curriculum reform. Liberal Studies are composed of modules focusing on topics drawn from three areas: 1. Personal Development, 2. Society and Culture, and 3. Science, Technology, and the Environment. The students in the study were taught by six different teachers who collaborated in preparing the syllabus and the assessment rubric. Students received equivalent instruction and assessment for each topic as the same teachers taught the same topics to all four classes.

### 4.2 Task Description and Online Assessment

The study focused on how students used online rubrics to assess the written arguments of peers. At the end of each semester teachers selected several written arguments, from earlier assignments and uploaded them to the online assessment platform (see Figure 1) for students to assess. Students evaluated 4 to 6 essays during each round of the assessment exercise. Students wrote arguments to support their positions and claims on issues pertaining to topics covered during the course, such as "*Do you agree with the statement that wealth is the only element affecting our quality of life*?" Teachers chose student essays that represented low, medium, and high levels of argumentation. Selected essays were accompanied by assessment rubrics. Teachers also uploaded their own grades and comments for selected essays for students to consult after they had finished assessing the same essay. Upon logging into the assessment area, students saw the essays to be evaluated without the name of author but with the grades and comments assigned by the teacher (Figure 1).
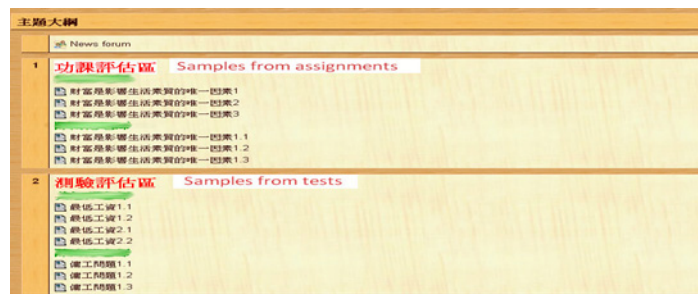


Figure 1. Screenshot of Peer assessment platform where students can download peers' sample work

Students had to select at least one essay on each assigned topic and assess it based on a rubric composed of four argument features: (a) claim, (b) reasoning, (c) evidence, and (d) application of knowledge. Each feature was associated with a 3-point scale: 0: poor, 1: fair, 2: good. Teachers provided detailed descriptions of what constituted a good, fair and poor claim, reasoning, evidence, and application of knowledge. For example, a good claim presented a "clearly stated and consistent point of view on the argument" while a poor claim presented a "vaguely stated or inconsistent point of view on the argument". Students assessed arguments by assigning them values for the four features. The online assessment interface, shown in Figure 2a, has two areas: a rubric area and a comments area. As students assessed each feature the color of the cell associated that feature changed according to the assigned value. The system calculated total scores for completed assessments (See Figure 2b). Students could also make comments on written arguments in the comments area.

After saving and submitting their assessments, a button appeared prompting students to compare their assessment with that of their teacher (see Figure 2c). Figure 2d shows a comparison between the assessment of a

student and that of a teacher. We hypothesized that in comparing their assessments with those of their teachers students would reflect on their own assessments and deepen their understanding of rubric criteria. The process of comparison could enhance the ability of students to construct and assess their own written arguments more competently.
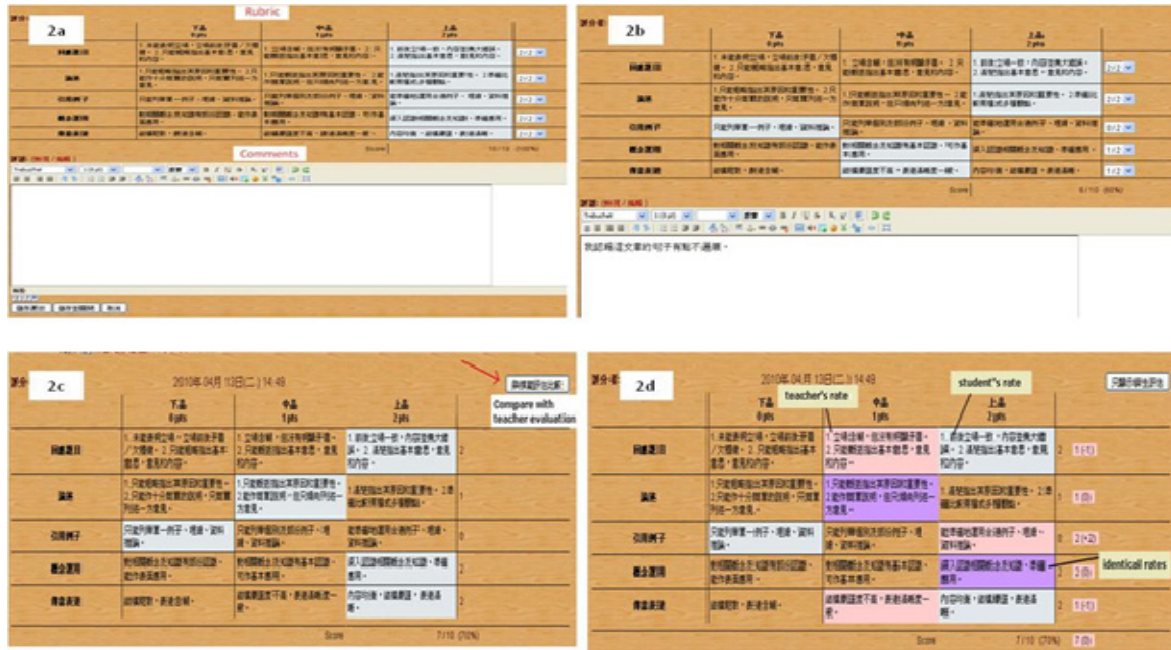


Figure 2. Screenshot of peer assessment platform where students can assess and compare the work

## 5. Data Sources and Analysis

Three types of data were collected for analysis: (a) first semester assignment scores and final exam grades; (b) self report surveys of peer assessment experiences and (c) online assessment activities.

Final exam grades were collected at the end of the school year and provided a holistic measure of students' argumentation skills. The final exam contained several short essay questions which teachers graded with rubrics resembling those used for regular assignments throughout the course. Students' overall scores on assignments in Semester 1, the semester preceding the peer assessment activity served as control variables.

Students completed a 10-item, on-line self-report survey in class at the end of Semester 2 to gain a better understanding of their online behavior and reflections. Items 1 and 2 dealt with how they assessed the work of peers (how they chose sample essays and whether they compared their assessments with those of their teachers). Items 3-9 dealt with two dimensions of online assessment: "assessment for reflection" and "assessment for learning". The internal reliabilities of "assessment for reflection" and "assessment for learning" were .681 and .748 respectively. Item 10 dealt with the usability of the online assessment system. Items 3-10 were associated with 5-point Likert scales where 1 stood for "strongly disagree" and 5 for "strongly agree".

Peer assessment activity data was collected and calculated for the whole year and consisted of the number and types of student comments and student-teacher agreements on the four features of arguments.

The database of the online assessment system recorded students' scores and comments for each written argument. Raw data was exported and compiled in excel files. Student- and teacher-assigned scores were compared. Same and different scores were coded "1" and "0" respectively. The total number of same scores was calculated for each of the four rubric features. Comments were coded based on our earlier work on peer assessment (author, 2011) which we adapted from Nelson and Schunn (2009), and Tseng and Tsai (2007). Thus, comments were first coded as affective and/or cognitive. Although, affective comments were coded as positive (e.g. "very good") or negative (e.g. "badly written") there were so few of each that we did not differentiate them in the study. We categorized cognitive comments as (a) identify problem; (b) suggestion; (c) explanation; and (d) comment on language. The author and a research assistant coded the cognitive comments independently with an inter-rater

reliability of .83. Comments that were neither cognitive nor affective (e.g. "can't read your project, cannot comment") were classified as 'other' and were later excluded as there were very few.

Table 1. Coding schema of peer feedback

| Categories | Definition | Example |
|---|---|---|
| **Cognitive** | | |
| Identifying problems | Addressing specific issue or dimension of the argument essay | *Claims are not very clear in the argument* |
| Suggestions | A method is suggested to deal with the problem | *You should enrich your argument with more examples and evidence* |
| Explanation | Explanation or elaboration on the problems identified or suggestions provided | *You did not explain why minimum wages can bring benefits to low-income workers* |
| Language | Comments addressing the writing in general | *Your writing is not quite clear.* |
| **Emotion** | | |
| Negative | Give criticism | *Lousy work!* |
| Positive | Praise the work | *Good!* |
| **Others** | | *You should work harder!* |

Statistical analyses were used to investigate the influences of the perceived effects of online assessment and actual online assessment activities on final exam scores. A multiple regression analysis (enter method) was conducted. The control variable was assignment performance in semester 1. We ran Pearson partial correlations to identify variables for inclusion in the regression model when controlled with Assignment scores (Table 2). The regression model dependent variable was the final exam score, and the independent variables were those significant in the partial regression table which included Example-match, Iden-problem, and Survey reflection.

Table 2. Correlation matrix among Final Exam score and correlates

| | Measure | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | Exam | | | | | |
| 2 | Example_match | .178* | | | | |
| 3 | Survey_Reflection | .228** | -.066 | | | |
| 4 | Iden_problem | .208* | .080 | .110 | | |
| 5 | Emotion | -.185* | -.044 | .029 | -.119 | |

* $p < 0.05$; ** $p < 0.01$.

Selected variables from partial correlations were checked for abnormalities in terms to multicollinearity and distribution. Since all variables entered into the regression model were continuous, relationships between them (multicollinearity) were investigated by examining Pearson partial correlations between pairs of variables, in which assignment score was controlled. No interrelations were found between predictors (see Table 2). In addition, all the variables had a normal distribution except "Emotion", which had a positively skewed (skewness = 5.25) and peaked distribution (kurtosis = 30.61). Data for "Emotion" was excluded from analysis because it remained abnormal even after the application of Log transformation.

## 6. Results

### 6.1 Summary Statistics

Table 3 presents a descriptive summary of final grades, assignment scores, total number of same assessments across rubric dimensions, number of different types of comments, and survey scores for the two factors.

Table 3. Descriptive analysis of learning performance, assessment activities, and survey report.

|  | Min. | Max. | M | SD |
|---|---|---|---|---|
| *Learning performance* |  |  |  |  |
| Assignment | 17.16 | 73.59 | 45.66 | 11.81 |
| Exam | 10.20 | 66.30 | 36.12 | 11.46 |
| *Report on web-based assessment* |  |  |  |  |
| Claim_match | 2.00 | 10.00 | 5.10 | 1.75 |
| Reasoning_match | 0.00 | 7.00 | 2.90 | 1.36 |
| Example_match | 0.00 | 8.00 | 2.69 | 1.38 |
| Knowledge_match | 0.00 | 7.00 | 2.98 | 1.25 |
| *Survey* |  |  |  |  |
| Survey_reflection | 1.25 | 5.00 | 3.58 | 0.70 |
| Survey_learning | 1.00 | 5.00 | 3.51 | 0.80 |
| *No. of comments* |  |  |  |  |
| Iden_problem | 0.00 | 8.00 | 3.35 | 1.58 |
| Suggestion | 0.00 | 4.00 | 0.53 | 0.97 |
| Language | 0.00 | 4.00 | 0.85 | 0.95 |
| Explanation | 0.00 | 7.00 | 3.14 | 1.40 |
| Emotion | 0.00 | 4.00 | 0.15 | 0.59 |

Most students chose assessment papers randomly (57%) and some chose purposely on different levels (21%). Fewer students chose good (17%) or poor articles (4%). Most compared their assessments with those of their teacher (90%) while a few did not (10%). Most students found the assessment system easy to use (M=3.68 on a 5-point Likert-scale where 5 meant very easy to use).

Although gender composition was unbalanced, there were no significant gender differences with respect to learning performance and assessment activities. An independent sample *t* test showed no significant difference between male and female participants on final exam scores, number of comments, and teacher-student assessment agreements.

### 6.2 Prediction for Final Exam Scores

The multiple squared correlation coefficient was 0.43, indicating that approximately 43% of the variability in the final exam was accounted for by assignment scores and the three predictors ("Example_match", "Survey_reflection", and "Iden_problem"). The control variable (Assignment) on its own was a significant predictor of Exam ($R^2 = 0.35$, $t = 8.08$, $p < 0.00$). When the three predictors were added to the model, there was a significant change of variance ($\Delta R^2 = 0.076$, $p < 0.01$). All the predictors were significant or marginal significant predictors (Example_match: $t = 2.02$, $p < 0.05$; Survey_reflection: $t = 2.51$, $p < 0.05$; Iden_problem: $t = 1.93$, $p = 0.057$) to Exam, controlling for Assignment. The strength of prediction of Assignment remained approximately the same after the three predictors were added ($t = 7.18$, $p < 0.00$). According to the standardized $\beta$ of the predictors, "Example_match", "Survey_reflection", and "Iden_problem" all predicted Exam positively.

Table 4. Regression model

|  | $R^2$ | $R^2_{adj}$ | $\Delta R^2$ | $\Delta F$ | Standardized $\beta$ | $t$ |
|---|---|---|---|---|---|---|
| Step 1 | .35 | .35*** | .35*** | 65.25 |  |  |
| Assignment |  |  |  |  | .60*** | 8.08 |
| Step 2 | .43 | .41*** | .076*** | 5.16 |  |  |
| Assignment |  |  |  |  | .52*** | 7.18 |
| Example_match |  |  |  |  | .14* | 2.04 |
| Survey_reflection |  |  |  |  | .18** | 2.51 |
| Iden_problem |  |  |  |  | .138† | 1.93 |

†$p < 0.1$, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.000$.

In this case, "Example_match", "survey_reflection", and "iden_problem" predict Exam positively. Overall, the three predictors accounted for 7.6% of the variance of Exam ($\Delta R2 = 0.076$, $p < 0.01$).

## 7. Discussion

This study examined whether students wrote better arguments after using rubrics to assess the written arguments of peers and then reflecting on and comparing their assessments with equivalent assessments by their teachers. Students' reflections on peer assessments, assessment agreement on evidence, and number of comments on identifying problems were found to significantly predict exam scores and to account for about 7.6% change of variance in the final exam after controlling for the effects of prior knowledge.

Of the four argument features, only evidence was a significant predictor for final exam scores. Student-teacher agreements on evidence significantly influenced exam performance. Evidence assessed whether students gave "single", "multiple but partial" or "sufficient" examples. Students with high assessment agreement for evidence did better in the final exam demonstrating that being able to competently assess the quality evidence in the arguments of peers was important in determining their ability to write good arguments. This implied that the ability to provide evidence for claims was the most important factor in being able to write good arguments and that when this feature is embedded in assessment rubrics, students were better able to judge the quality of arguments which in turn lead to better learning.

The non-significance of the other three features: claim, reasoning and application of knowledge, implies that argument skills do not develop all at once and that development may start with the ability to construct evidence followed later by the development of the ability to formulate claims, engage in reasoning, and achieve conceptual understandings. However, due to limits on the exercise of peer assessment, these other features did not positively affect learning performance. The findings suggest that the possibility that argument features develop sequentially should be considered when designing argument tasks. These findings further indicate that the ability to assess arguments and the ability to write arguments are not the same. Students may have difficulty transferring their skills in assessing arguments to writing argument.

Among the comments, identifying problems was a marginally significant predictor. More comments on identifying problems lead to better exam performance. This finding was consistent with our earlier work on the different effects of online peer assessment on assessors and assessees (Author, 2011). For instance, in providing cognitive comments such as identifying problems to peers, the assessor or person giving the comment benefited more that the assessee or person receiving it. Descriptive analysis revealed that this was the most frequent type of comment. Though students provide about the same number of Explanation comments and Identifying problems comments, the former was not a significant predictor for final exam scores. This could be because the Explanation comments were not constructive. Further, students offered few Suggestions perhaps because formulating solutions to problem draws on higher cognitive abilities. Thus, it appears that identifying problems, giving explanations, and suggesting solutions involve different levels of cognitive capability. Students also provided few language and emotion comments perhaps because the assessees simply did not receive the comments. Thus, assessors may not have been motivated to provide suggestions, comments on language issues or emotional issues.

Survey analysis revealed that Assessment-for-reflection was a significant predictor on exam performance with those who engaged in more online peer assessment doing better on the final exam. This study was designed to get students to compare how they assessed the arguments of peers with how their teachers assessed them. The idea being that in doing so students would reflect on the rubrics and thus on what constitutes a good argument.

The results are consistent with our hypothesis that students who reflected more on their performance did better on the final exam.

This study used three methods of assessment: rubric-based, feedback, and on-line technology. It didn't seek to determine whether there were performance differences between those engaging in and those not engaging in rubric-based assessment as in earlier research (Hughes, 1995). Rather, we implemented a model of argumentation in the assessment rubric and tried to identify how assessing different features of arguments might influence the ability of students to write arguments. The rubric induced students to focus on features of arguments and provided instruction on the cognitive skills needed to assess the written arguments of peers. On-line assessment provided students with tools for visually representing and sharing the procedures and results of assessing arguments so as to concentrate attention and induce reflection. Results suggested that argumentation assessment skills can be improved by involving students in peer assessment activities. The rubrics provided students with clear guidance on assessing different argument features which in turn lead them to improve their ability to write arguments. The task of using rubrics to assess peer arguments helped students differentiate well-constructed arguments from poorly-constructed ones. To apply the rubric students had to understand the criteria specified in it and how to apply them to peer arguments. In so doing students become more aware of the characteristics of well- and poorly-constructed arguments which lead them to use the same criteria more reflectively and attentively in assessing their own written arguments. The effects of rubric-based assessment on learning are indirectly reflected in students' self-reported surveys.

Students who encountered disagreements between their assessments and those of their teachers were lead to reflect on why and may have induced them to review the assignment. Assessment activities also helped students to understand the features of arguments better by sharpening their critical thinking skills. Peer assessment can promote self assessment (Liu & Carless, 2006). Students can gain insights into their own work by judging and critiquing the work of peers (Bostock, 2000). Learners developed clearer and deeper understandings of task and argument dimensions by critically judging and commenting on the quality of peer written argument than by simply focusing on their own written arguments.

## 8. Conclusions and Future Directions

Argumentation involves different types of skills and students may fail to develop these skills in a balanced fashion, especially when it comes to evaluating the written arguments of peers. Judging the arguments of peers based on different argument criteria can help students develop a better understanding of the structure and quality of written arguments which can in turn help them to write better arguments. The development of argumentation skills can be facilitated by involving students in peer assessment activities. Since students perceived that they benefited more from assessing poor quality written arguments, they should be given more opportunity to do so, but with clearly stipulated rubrics and guidelines.

This study only investigated the effects of peer assessment on assessors due to the nature of assessment task. It would be interesting to see if assessees also benefited from such activities. For instance, how would assessees interpret inconsistencies between how peers and teachers assessed of their work? How would they interpret comments from peers on their written arguments? Would they able to integrate peer comments and revise their arguments accordingly? These issues should be investigated in future studies.

## 9. Limitations

One hundred and twenty one students were from four different classes. Since this sample was chosen due to convenience, the generalization of the results is constrained.   Stronger and more robust results will be possible via a random sampling procedure.

Considering the sample size, we did not examine the differences of argumentation skills among these four areas. By subdividing the groups, the sample size in each category would become smaller and the conclusions would be weaker. We believe that given sufficient sample size in each category of the four areas, it would be very interesting to examine argumentation differences.

## References

Aberdein, A. (2005). The uses of argument in mathematics. *Argumentation, 19*(3), 287-301. http://dx.doi.org/10.1007/s10503-005-4417-8

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13-18.

Asterhan, C., & Schwarz, B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99*(3), 626-639. http://dx.doi.org/10.1037/0022-0663.99.3.626

Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 1-64). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bostock, S. (2000). Student peer assessment Keele University, Learning Technology Web site. Retrieved May 5, 2010, from http://www.palatine.ac.uk/files/994.pdf

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604. http://dx.doi.org/10.1207/s15516709cog2404_2

Chang, S. N., & Chiu, M. H. (2008). Lakatos' scientific research programmes as a framework for analysing informal argumentation about socio-scientific issues. *International Journal of Science Education, 30*(13), 1753-1773. http://dx.doi.org/10.1080/09500690701534582

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4,* 45-69. http://dx.doi.org/10.1016/0959-4752(94)90017-5

Chiu, M. M. (2008). Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemporary Educational Psychology, 33,* 382-402. http://dx.doi.org/10.1016/j.cedpsych.2008.05.001

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84,* 287-312. http://dx.doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.3.CO;2-1

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38,* 39-72. http://dx.doi.org/10.1080/03057260208560187

Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In S. Erduran, & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education* (Vol. 35, pp. 47-69): Springer Netherlands.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment & Evaluation in Higher Education, 11*(2), 146-166. http://dx.doi.org/10.1080/0260293860110206

Falchikov, N., & Blythman, M. (2001). Learning together: Peer tutoring in higher education (1st ed.). New York: Routledge.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2009). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304-315. http://dx.doi.org/10.1016/j.learninstruc.2009.08.007

Goldstein, M., Crowell, A., & Kuhn, D. (2009). What constitutes skilled argumentation and how does it develop? [argument; argumentation; development; metacognition; causal reasoning]. *Informal Logic, 29*(4), 379-395.

Hagler, D. A., & Brem, S. K. (2008). Reaching agreement: The structure & pragmatics of critical care nurses' informal argument. *Contemporary Educational Psychology, 33*(3), 403-424. http://dx.doi.org/10.1016/j.cedpsych.2008.05.002

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review, 114*(3), 704-732. http://dx.doi.org/10.1037/0033-295X.114.3.704

Hughes, I. E. (1995). Peer assessment. *Capability, 1,* 39-43.

Jamaludin, A., Chee, Y. S., & Ho, C. M. L. (2009). Fostering argumentative knowledge construction through enactive role play in Second Life. *Computers & Education, 53*(2), 317-329. http://dx.doi.org/10.1016/j.compedu.2009.02.009

Jonassen, D., & Kim, B. (2010). Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology Research and Development, 58*(4), 439-457. http://dx.doi.org/10.1007/s11423-009-9143-8

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144. http://dx.doi.org/10.1016/j.edurev.2007.05.002

Karacapilidis, N., & Papadias, D. (2001). Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems, 26*(4), 259-277. http://dx.doi.org/10.1016/S0306-4379(01)00020-5

Knudson, R. E. (1991). Effects of instructional strategies, grade and sex on students' persuasive writing. *Journal of Experimental Education, 59*(2), 141-152.

Knudson, R. E. (1992). The development of written argumentation: An analysis and comparison of argumentative writing at four grade levels. *Child Study Journal, 22*(3), 167-184.

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1-8. http://dx.doi.org/10.1111/1467-9280.00302

Kuhn, D. (2005). *Education for thinking*. Cambridge, Massachusetts: Harvard University Press.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260. http://dx.doi.org/10.1111/1467-8624.00605

Kwok, R. C. W., & Ma, J. (1999). Use of a group support system for collaborative assessment. *Computers & Education, 32*(2), 109-125. http://dx.doi.org/10.1016/S0360-1315(98)00059-1

Lajoie, S. P., Greer, J. E., Munsie, S. D., Wilkie, T. V., Guerrera, C., & Aleong, P. (1995). Establishing an argumentation environment to foster scientific reasoning with Bio-World. In D. Jonassen, & G. McCalla (Eds.), *Proceedings of the International Conference on Computers in Education* (pp. 89-96). Charlottesville, VA: Association for the Advancement of Computing in Education.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education, 77*(4), 339-366. http://dx.doi.org/10.3200/JEXE.77.4.339-366

Li, D. D., & Lim, C. P. (2008). Scaffolding online historical inquiry tasks: A case study of two secondary school classrooms. *Computers & Education, 50*(4), 1394-1410. http://dx.doi.org/10.1016/j.compedu.2006.12.013

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 Technology, 41*(3), 525–536. http://dx.doi.org/10.1111/j.1467-8535.2009.00968.x

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*(4), 420-432. http://dx.doi.org/10.1046/j.0266-4909.2001.00198.x

Liu, E. Z. F., Lin, S. S. J., Chiu, C. H., & Yuan, S. M. (2001). Web-based peer review: The learner as both adapter and reviewer. *IEEE Transactions on Education, 44*(3), 246-251.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279 -290. http://dx.doi.org/10.1080/13562510600680582

Lu, J., & Lajoie, S. P. (2008). Supporting medical decision making with argumentation tools. *Contemporary Educational Psychology, 33,* 425-442. http://dx.doi.org/10.1016/j.cedpsych.2008.05.005

Moskal, B. M. (2000). Scoring rubrics: What, when and how. Retrieved from http://pareonline.net/getvn.asp?v=7&n=3

National Assessment of Educational Progress. (1998). NAEP 1998 Writing Report Card: Findings from the National Assessment of Educational Progress. Princeton, NJ: Educational Testing Service.

National Science Board. (2006). America's pressing challenge-building a stronger foundation. Retrieved Sept. 20, 2010, from http://www.nsf.gov/statistics/nsb0602/nsb0602.pdf

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. http://dx.doi.org/10.1007/s11251-008-9053-x

Nussbaum, E. M. (2002). Scaffolding argumentation in the social studies classroom. *The Social Studies, 93*(2), 79-83. http://dx.doi.org/10.1080/00377990209599887

Nussbaum, E., Winsor, D., Aqui, Y., & Poliquin, A. (2007). Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 479-500.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*(3), 239-250. http://dx.doi.org/10.1080/0260293960210304

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 83-105). Hillsdale, NJ: Erlbaum.

Royer, J., Cisero, C., & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research, 63*(2), 201-243. http://dx.doi.org/10.3102/00346543063002201

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*(5), 513-536. http://dx.doi.org/10.1002/tea.20009

Stefani, L. A. J. (1994). Peer, self, and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75. http://dx.doi.org/10.1080/03075079412331382153

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-277. http://dx.doi.org/10.3102/00346543068003249

Topping, K. J., & Ehly, S. W. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation, 12*(2), 113-132. http://dx.doi.org/10.1207/S1532768XJEPC1202_03

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education, 25*(2), 149-169. http://dx.doi.org/10.1080/713611428

Toulmin, S. E. (1958). The use of argument. Cambridge: Cambridge University Press.

Tsai, C. C. (2009). Internet-based peer assessment in high school settings. In L. T. W. Hin, & R. Subramaniam (Eds.), *Handbook of research on new media literacy at the K-12 level: Issues and challenges* (pp. 743-754). Hershey, PA: Information Science Reference.

Tsai, C. C., Lin, S. S. J., & Yuan, S. M. (2001). Developing science activities through a networked peer assessment system. *Computers & Education, 38*(1-3), 241-252. http://dx.doi.org/10.1016/S0360-1315(01)00069-0

Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*(4), 1161-1174. http://dx.doi.org/10.1016/j.compedu.2006.01.007

Van Lehn, K. A., Chi, M. T. H., Baggett, W., & Murray, R. C. (1995). Progress report: Towards a theory of learning during tutoring. Pittsburgh, PA: Learning Research and Development Centre, University of Pittsburgh.

Walton, D. N. (1996). Argumentation schemes for presumptive reasoning. Mahwah, NJ: Lawrence Erlbaum Associates.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301-311. http://dx.doi.org/10.1037//0022-0663.91.2.301

Wu, Y. T., & Tsai, C. C. (2007). High School Students' Informal Reasoning on a Socio-scientific Issue: Qualitative and quantitative analyses. *International Journal of Science Education, 29*(9), 1163-1187. http://dx.doi.org/10.1080/09500690601083375

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*(3-4), 186-193. http://dx.doi.org/10.1016/j.iheduc.2008.06.005

Yang, Y. F. (2010). A reciprocal peer review system to support college students' writing. *British Journal of Educational Technology, 42*(4). http://dx.doi.org/10.1111/j.1467-8535.2010.01059.x