A Simulated Peer-Assessment Approach to Improve Students' Performance in Numerical Problem Solving Questions in High School Biology

Fraser J. Scott

School of Education, University of Strathclyde, Glasgow, Scotland

E-mail: fraser.j.scott@strath.ac.uk.

Dr Fraser Scott obtained his PhD in drug design from the University of Strathclyde in 2012 under the direction of Prof. Colin Suckling. He subsequently trained as a teacher of chemistry, also at the University of Strathclyde, and currently teaches at a Scottish high school. His research interests lie in the area of student difficulties in mathematics and how these impinge on performance in the sciences.

A Simulated Peer-Assessment Approach to Improve Students' Performance in Numerical Problem Solving Questions in High School Biology

Basic mathematical fluency is a prerequisite for success in a wide array of areas in biology and it has been noted that many students are deficient in this skill. In this paper, the use of a *simulated* peer-assessment activity is investigated as a method to improve performance in numerical problem solving questions in high school biology. Additionally, the benefits of using *simulated*, rather than real, student's answers in peer-assessment is discussed. The study involved a small cohort of students who carried out a *simulated* peer-assessment as a classroom activity and their improvement in performance and attitude towards the activity was measured. The results demonstrate that a *simulated* peer-assessment activity is suitable as a replacement for standard peer-assessment and that students' attitudes favour the simulated approach.

Keywords: peer-assessment; self-assessment; numerical problem solving; mathematical deficiency; high school biology

Introduction

Numerical Problem Solving in Biology Education Review

Problem Solving

"Problem solving" is a term that is encountered frequently in biology education research, and, more broadly, science education research. It is a fundamental skill that is necessary both to enable the learning journey, from novice to expert, and to allow an expert to operate in their chosen field (American Association for the Advancement of Science, 2011; National Academy of Science, 2011). Problem solving can be defined as the application of one, or more, operations in order to transfer the given state of a system to a goal state (Newell & Simon, 1972; Dunbar, 1998). This definition is rather general but necessarily so as the literature on problem solving is vast. Despite this, two key features of numerical problem solving, or problem solving as a whole, can be identified as being particularly significant in the research presented in this paper: level appropriateness; and, novelty.

In determining whether or not a question can be described as a "problem", or one requiring "problem solving" skills, one cannot examine the question in isolation from its intended target. The question must be developmentally appropriate for the students in order for it to be considered a problem solving question (Lesh & Zawojewski, 2007; Piaget & Inhelder, 1975). For example, as students are progressing through their Biology education they are likely to be progressing through their mathematical education too and thus, at different points during this progression, their mathematical fluency will differ. A numerical biology question, requiring a particular mathematical operation, might be routine for a student late in their education whilst a student at an earlier stage in their education may find the same operation, and overall question, much more problematic. A question can simultaneously be both a problem solving question and a routine exercise, depending on the student.

Once a problem has been solved, it is no longer a problem and thus familiarity with a question influences its categorisation as a problem. This idea of novelty being necessary for problem solving was first discussed by Köhler in 1925 (Köhler, 1925). Since then, many researchers have echoed his thoughts (Polya, 1945 & 1962; Schoenfeld, 1985; National Council of Teachers of Mathematics, 2000). Novel questions require higher order thinking skills in order to reason and solve and the necessity of these skills implies that any operation that is performed automatically cannot be problem solving (Lester & Kehle, 2003; Resnick & Ford, 1981). Thus, a question which can be solved by an algorithm alone, applied without thought, does not constitute a problem solving question. Rather, one would consider this question to be a routine exercise.

Numerical Proficiency in Science

A sound understanding of basic mathematical principals is fundamental to be able to understand most scientific phenomena. High school students' lack of this basic mathematical understanding, and its impact on science, has been the focus of much media attention of late (Royal Society of Chemistry, 2009a, 2012a-b); moreover, a recent report from SCORE (Science Community Representing Education, 2010), a collaboration of science organisations with an interest in education, has highlighted that a significant proportion of the mathematical requirements of high school science courses are not assessed. The importance of identifying and improving mathematical inadequacies has led to the development of several initiatives by members of SCORE to address this concern (Royal Society of Chemistry, 2009b-c).

The problem of numerical problem solving has been well documented within science education research. This is particularly the case for physics and chemistry education research; the literature in biology education research does document numerical proficiency as an issue but is limited by comparison. Researchers in the field of physics education have investigated numerical problem solving for at least three decades (Hsu, Brewe, Foster & Harper, 2004). This has focused on the definition of problems in physics (Heller and Reif, 1984), the investigation of how students solve problems (Larkin, McDermott, Simon & Simon, 1980), and the development of effective pedagogy to effect deep learning (Heller and Hollabaugh, 1992; Pawl, Barrantes, & Pritchard, 2009). It has been argued that the use of mathematics in physics is more complex than the application of rules and calculations that students are taught in a mathematics class (Bing & Redish, 2009). Apata (2013) has stated that a sound understanding of basic mathematical skills is necessary for students to engage with numerical problem solving in physics. Apata (2013) has also commented that many students do not possess these necessary mathematical skills (WACE, 2004). The use of mathematical language in physics differs in many ways from its use in mathematics (Redish, 2006) and this could be a reason for the lack of transfer of numeracy skills from a mathematics to a physics setting (Bing, 2008).

The necessity for students to be numerically proficient in order for them to engage with large parts of chemistry curricula, in particular chemical calculations, has been well documented within the field of chemistry education research. As far back as 40 years ago it was thought necessary to prepare ‘‘Mathematical Readiness’’ tests to equip students with a fundamental set of mathematical skills in order for them to perform and understand chemical calculations. Weisman (1981) identified key areas of High School chemistry courses that require mathematical skills, including: stoichiometry problems; Avogadro’s number; and, balancing equations. Denny (1971) analysed the chemical calculations that High School students were expected to be able to perform and suggested ten fundamental mathematical skills: computation; use of parentheses; signed number usage; use and manipulation of fractions; use of decimals; use of exponents, manipulation of numbers with exponents and logarithmic equivalence; use of percentages; manipulation of one-variable equations; use of ratio and proportion; and, producing and interpreting x, y graphs. Proficiency in these skills are not just necessary within chemistry but all of science education.

Lazonby, Morris and Waddington (1982) have investigated numerical problem solving in the context of chemistry calculations and concluded that the inability to carry out the sequential application of fundamental mathematical operations was a significant factor to students’ lack of success. To circumvent this problem, an algorithmic approach to obtaining a solution is widely used, such as in dimensional analysis. However, it has been suggested that although these methods may generate the correct answer, the students employing these methods do not truly possess understanding (Lythcott, 1990).

DeLorenzo (1994) developed a more language rich version of dimensional analysis, with the inclusion of verbs and nouns in the description of the chemical system, which may impart a greater degree of understanding compared to the classical approach.

Student mathematical proficiency, or lack thereof, has been previously commented on within the field of biology education over the past decade (Gross, 2000; Bialek and Botstein, 2004). O'Shea (2003) found that Irish students perform particularly poorly on biology questions that require mathematical tasks presented in a non-routine format. A study of Australian nursing students by Eastwood *et al.* (2011) highlighted basic mathematical errors in their calculation of drug concentrations, a skill commonly encountered in biology orientated jobs. A decade-long survey of biology students, focusing in the area of plant physiology, revealed persistent weaknesses in students' abilities to answer quantitative questions (Llamas *et al.*, 2012).

Gross (2000) investigated the effect of mathematics on biology and discovered that these courses are often taught almost independently of each other at university level. Gross asserts that this approach leaves students with domains of knowledge that do not overlap and students find it difficult to apply knowledge from one course to the other. Similarly, Hourighan and O'Donoghoue (2006) have found that students enter mathematically demanding courses at university with a distinct lack of the basic mathematical skills needed to succeed. Their subsequent investigation into this revealed, again, that mathematics is taught in isolation to the biology; students are not given the opportunity to explore the mathematical ideas in a contextualised setting; and that this method of teaching promotes a 'learned helplessness' within the student body. Bialek and Botstein (2004) came to a similar conclusion: they argued that biological sciences are becoming too complex to begin the learning of the interconnectedness across various fields of study at a late stage. They assert that a more integrated approach is required early on in students' education.

It is clear that students' lack of mathematical ability is a concern for biology education, and indeed the whole of science education; however, there is very little research to indicate who should be solving this 'mathematics problem': should it be the mathematics teachers or the biology teachers? Irrespective of the origin of the 'mathematics problem' there are students in biology classes that have difficulties with mathematics and biology educators have a responsibility to assist.

Peer-Assessment Review

Topping (1998) defined peer-assessment as the process of evaluating the quality or success of the outcomes of a peer or peers which is followed by the provision of feedback (Van Den Berg, Admiraal & Pilot, 2006). To the peer being assessed, the provision of feedback is of benefit to allow them the opportunity to improve; however, this is not necessarily the most useful benefit of peer-assessment. The evaluation process that a student must engage with in order to provide feedback is arguably the most important aspect of peer assessment (Ljungman & Silen, 2008; Topping, 2005). This is so because it affords the student the opportunity to practice the skills necessary to evaluate their own work and thus peer-assessment serves to improve skills in self-assessment (Anderson, Howe, Soden, Halliday, & Low, 2001; Topping, 2005).

O'Donovan, Price and Rust (2004) have remarked that when students are fully engaged in the marking process, they can transform the assessment of learning into assessment for learning and thereby obtain a skill to enhance their learning tool. Both peer and self-assessment techniques enhance autonomy in students' learning and help to foster a metacognitive skill set (Brown and Knight, 1994, Elwood and Klenowski, 2002). Candy *et al.* (1994) highlight the importance of peer and self-assessment in their finding that identification of educational needs is fundamental to successful lifelong learning. Harlen (2007) too describes the requirement of responsibility a student has for their own learning as it can benefit, not just the student's life after school but, society as a whole.

Students who are instructed through peer and self-assessment methods have been shown to be cognisant of the benefits. The views of primary school students to peer-assessment were investigated by Bryant and Carless (2009) and a very positive reception was displayed when the students were given the facility to learn from each other whilst taking responsibility for their own work resulted. Bryant and Carless' study also discovered that some students were acutely aware of the advantages that peer-assessment brought as they could identify errors in their work and thereby avoid them in future. Peterson and Irving (2008) discovered that secondary school students believed that feedback provided through peer-assessment was motivational and encouraged them to seek out solutions in amelioration of their errors. The broader benefits of peer and self-assessment are not lost on high school students; White (2009) reports that the opportunity to enhance skills which are helpful for their future career was significant motivator.

Boud (1995) and McDonald & Boud (2003, p 210) have argued that the skills developed through peer and self-assessment are of great importance throughout all stages of education. Furthermore, a thorough review of the literature resulted in Black and William (1998) concluding that self-assessment is "not an interesting option or luxury: it has to be seen as essential" (p54-55). These positive views on peer and self-assessment are validated in practice: Rust *et al.* (2003) and O'Donovan *et al.* (2004) demonstrated that participation in a peer-assessment program at the beginning of a course of study resulted in an enhancement in performance over those students who did not participate.

Whilst the advantages of peer-assessment are significant there are a number of issues that must be considered in its opposition. Wen and Tsai (2006) found that university students' attitudes towards peer-assessment were generally positive; however, there was a lack of self-confidence in their ability to mark their classmates' work and, reciprocally, they were apprehensive about peer criticism. Karaca (2009) encountered similar results when carrying out a study into teacher trainee's opinions of peer assessment. It was suggested that students find it difficult to evaluate their peers' work effectively, leading to the provision of deleterious feedback. Karaca also found that students' evaluations could be influenced by their social relationships with their peers: friendly students were prone to providing overly positive feedback; contrariwise, rivalrous students increased their provision of negative feedback. Ballantyne, Hughes and Mylonas (2002) have also reported that students can be apprehensive about the time consuming nature of peer-assessment.

Bostock (2000) and White (2009) assert that validity and reliability of peer-assessment may be an issue as the feedback provided may not be accurate or valuable and even that some students may not take the assessment process seriously. Additionally, they concur with Karaca (2009) in that students are not necessarily skilled enough to effectively evaluate each other and that some students may be influenced by social standing. Bostock (2000) and White (2009) further note that an absence of teacher input to the evaluation process could lead to the provision of mis-information.

Bryant and Carless (2009) have found that a student's perception of peer-assessment can differ depending on the relative language proficiency of themselves and their peer. Those who were assessed by peers of superior language proficiency indicated that it was difficult to assess their peer's work due to the ability difference; contrariwise, more able students found that their peers could not provide actionable feedback. Bryant and Carless asserted that teachers were more reliable source of feedback.

Peer assessment in Mathematics Learning

The field of mathematics education has explored the utility of peer-assessment strategies in a mathematics classroom. Kroeger (2006) has asserted that peer-assisted learning strategies are able to positively influence high school students' attitudes to difficult areas of the mathematics curriculum. Chukwuyenum and Adeleye (2013) demonstrated that students' were able to enhance their own performance in high school mathematics after engaging in a peer-assessment activity. In a detailed study into the effects of both peer and self-assessment in a high school mathematics classroom, Hammonds (2013) has demonstrated that positive effects are observed in motivation and mathematical performance; however, these are very much more pronounced in low and mid-level learners. These findings from the mathematics education literature suggest that peer-assessment may be an appropriate tool to assist with numerical problem solving questions in a biology setting.

Simulated Peer-Assessment

In an effort to reduce the negative aspects inherent in peer-assessment, the author has previously investigated the use of a *simulated* peer-assessment method (Scott, 2014). In this variation of peer-assessment, students were provided with an example of an incorrect solution to a problem and their goal was to analyse this solution, identify errors and provide feedback. Through removing the students' own work from the assessment proceedings, it was thought that the negative aspects such as apprehension about being criticised; provision of deleterious feedback; influence of social relationships; and, differing peer abilities could be removed. The evaluation process, which is a critical aspect of peer-assessment, still remains in the simulated peer-assessment model and hence it was thought that this method would not diminish the positive aspects of peer-assessment.

The author's previous study investigated the utility of a *simulated* peer-assessment approach in the context of high school chemistry calculations. After participating in the simulated peer-assessment activity, students were able to avoid common sources of error and improve their performance as measured by a coupled pre-test and post-test. *Simulated* peer-assessment provided the greatest increase in performance to those of intermediate mathematical ability. It was thought that students of lower mathematical ability would benefit from direct instruction in simple mathematical skills to allow them to benefit from participation in this activity. The study also investigated what the difference in students' attitudes would be if the activity was not simulated but instead a straightforward peer-assessment. This study demonstrated that most students would be unhappy to have their peers assessing their work, perhaps due to the poorer quality of feedback that may be provided. Crucially, this study demonstrated that the use of simulated work does not result in a negative attitude from the students undertaking the activity; it had been thought that there may have been a level of dissatisfaction from the students if they were not marking 'real' work. In this context, *simulated* peer-assessment provided a mechanism through which the benefits associated with peer-assessment could be provided without having to deal with the negative aspects such as poor quality of feedback or inconsistent feedback due to social networks within a class. The findings from this previous study into simulated peer-assessment can be summarised as follows:

(1) Simulated peer-assessment was able to improve students' performance in chemical calculations.

(2) Students with the greatest improvement were of intermediate mathematical ability.

(3) Students did not display a negative attitude towards the use of 'simulated' rather than 'real' solutions within the activity.

The Purpose of This Study

This study aims to investigate the method of utilising a *simulated* peer-assessment to improve high school students' performance in numerical problem solving questions in Biology. The numerical problem solving questions that are under investigation are those commonly encountered by candidates sitting the Higher Biology course of the Scottish education system. The author holds the view that the main factor affecting performance in these numerical questions is the poor mathematical ability of students; however, methods are sought to ameliorate this problem in the Biology classroom rather than that of Mathematics. In this study, a numerical problem solving based *simulated* peer-assessment activity is thought to be an appropriate method to improve students' performance. The author has recently investigated the utility of a *simulated* peer-assessment approach in the context of high school chemistry calculations (Scott, 2014) and this present study hence aims to replicate the investigation into *simulated* peer-assessment from a biology perspective.

Research Questions

This study used a mixed-methods design to investigate the usefulness of a *simulated* peer-assessment activity in improving students' ability in numerical problem solving questions in high school level biology. Three research questions were identified:

(1) To what extent does participation in the simulated peer-assessment activity increase students’ performance in numerical problem solving question in Biology?

(2) To what extent are students' attitudes different with simulated peer-assessment than with straightforward peer-assessment?

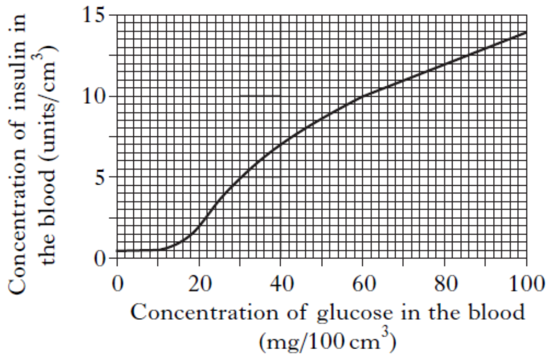
Research Methodology

Situational and Structural Analysis

In order to design a simulated peer-assessment activity in a biology setting it was first necessary to conduct a review of the types of question encountered by students following the Scottish higher biology course. This situational and structural analysis of the problem domain was carried out as follows: situational refers to the identification of biology contexts where numerical problem solving skills are used in the higher biology course and structural refers to the examination of the specific numerical skills that are required to solve such problems. Both of these stages involved discourse with multiple, practising high school biology teachers. This analysis identified four distinct question types and these are listed in table 1 along with an example of each. Each of these questions first involves extracting the relevant numerical details from either a graph, table or passage of information before the appropriate mathematical skills can be used to solve the question.

Table 1: Question types and examples.

Question Type	Example
Proportion	The graph below shows how the concentration of insulin in the blood varies with the concentration of glucose in the blood.



What total mass of glucose would be present at an insulin concentration of 10 units/cm3, in an individual with 5 litres of blood?

Percentage

The table shows the masses of various substances in the glomerular filtrate and in the urine over a period of 24 hours.

Substance	*Mass in glomerular filtrate* (g)	*Mass in urine* (g)
Sodium	600·0	6·0
Potassium	35·0	2·0
Uric acid	8·5	0·8
Calcium	5·0	0·2

What percentage of the total mass of substances found in the urine is potassium?

Ratio

Leaf rust is a fungus which grows when its spores land on leaves. The fungus spreads over leaf surfaces causing damage. Single leaves from four different species of cottonwood tree were sprayed with identical volumes of a suspension of rust fungus spores. After 3 days the percentage of leaf area with fungal growth was measured. The tannin content in these leaves was also measured.

Cottonwood species	*Percentage leaf area with fungal growth after 3 days*	*Tannin content in leaves* (mg per g of leaf dry mass)
Black	2·4	40·6
Eastern	11·4	3·9
Narrow-leafed	4·3	11·7
Swamp	3·2	15·6

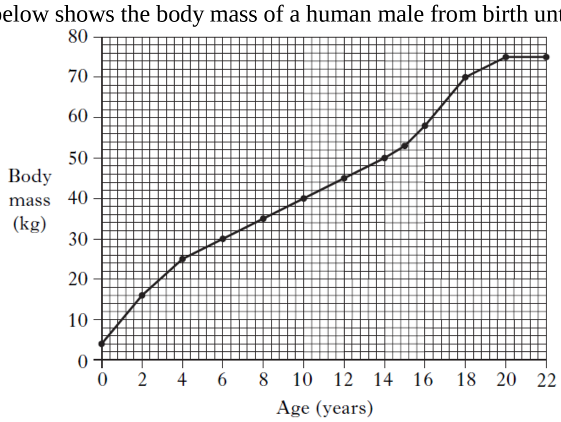
Express as the simplest whole number ratio, the tannin content in the leaves of the eastern cottonwood, narrow-leafed cottonwood and swamp cottonwood.

Percentage

Increase or

decrease

The graph below shows the body mass of a human male from birth until 22 years of age.



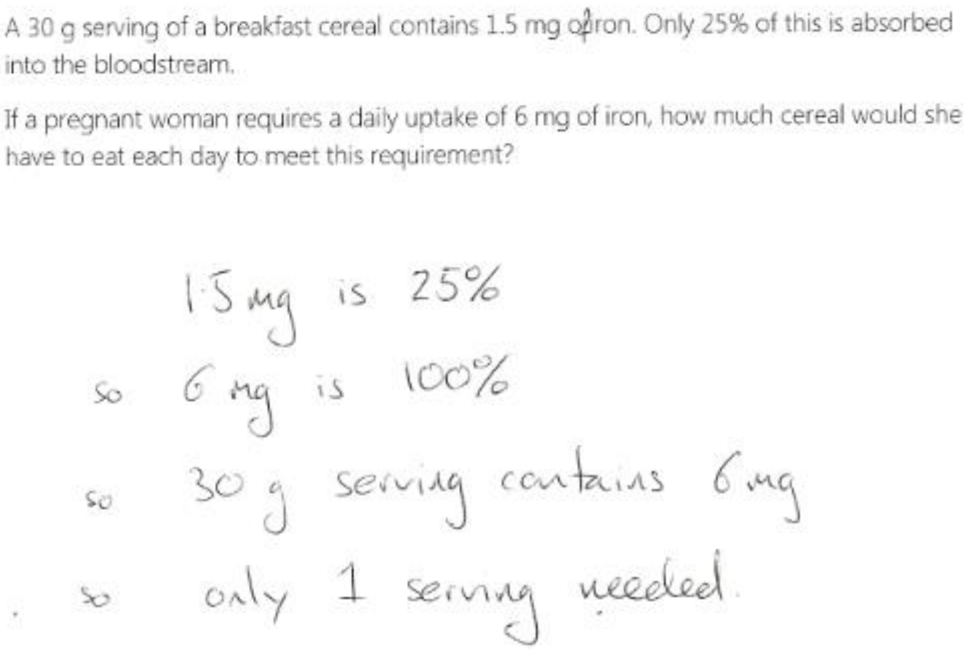
Calculate the percentage increase in body mass in the last 8 years of the study.

In carrying out the situational analysis of the four types of numerical problem solving questions it became important to rationalise the use of the term ‘problem’ in the context of these questions. To the more experienced scientist these questions would seem to be nothing more than simple ‘exercises’; however, it is thought that a number of factors contribute to their classification as ‘problems’ within this context. First, the contextualisation of the mathematics within a biology setting increases the difficulty of these questions and makes them more than routine mathematics exercises. Second, although these questions are part of the Higher biology curriculum, biology teachers view these as problem solving questions, and refer to them as such, so often do not ‘teach’ them like they would the rest of the curriculum. Biology teachers rely on students’ previous instruction in mathematics to solve these questions themselves. It is thus thought that the lack of familiarity students have with these questions in a biology setting also contributes to their elevation from ‘exercises’ to ‘problems’. The author does recognise that the classification of these questions as ‘problems’ or ‘exercises’ may be a debatable one.

The question types identified through the situational and structural analysis were then used to design the *simulated* peer-assessment activity, details of which are described in the next section.

Activity Design

This activity provides students with a series of questions as would be expected of those sitting the Scottish Higher Biology course and, along with each question, a sample solution is also provided (fig. 1). The solution, however, is incorrect and the aim of the activity is for students to identify any errors present, write a comment beside the solution to explain why there is an error (as if they were a teacher) and to provide a correct solution to the question.



A 30 g serving of a breakfast cereal contains 1.5 mg of iron. Only 25% of this is absorbed into the bloodstream.

If a pregnant woman requires a daily uptake of 6 mg of iron, how much cereal would she have to eat each day to meet this requirement?

1.5 mg is 25%

so 6 mg is 100%

so 30 g serving contains 6 mg

so only 1 serving needed.

Figure 1: Example of a simulated solution a student would analyse (error present).

There are twelve questions, and associated solutions, presented to the students during the simulated peer-assessment activity and these are divided into the four question types identified in the situational and structural analysis. Each type of question is a standard question that a Higher Biology student would be expected to carry out and the question types differ in the specific mathematical steps involved in calculating the correct answer.

Contextualised within the Biology setting, these questions first require either extraction of the relevant data from a short passage or from a graph or table before the mathematical computation can occur. The simulated solutions to the questions provided to the students include a spread of common errors that students often make, as identified through personal experience and through discourse with multiple, practising biology educators, and are spread amongst simple mathematical errors and errors in interpretation of the question. The simulated peer-assessment activity and a more detailed analysis of the activity design can be found as supplementary information.

When participating in the activity, the students were arranged into pairs or groups of three and were given the question sheets to work through. Group work is an important aspect of this activity as it promotes discussion and idea sharing within the group in order for them to discover the error and provide the correct solution. It is thought that the discussion throughout the activity should allow the students to become more aware of the common errors they are likely to make in their own work and thus prevent them from making them in the future. As the students carried out the activity, the teacher roamed the classroom to answer any queries that the students had.

Study Setting and Participants

This study was carried out in a high school in Scotland during the run up to the national exams thus none of the content of the activity would be considered new to the participating students. All four of the school's Higher Biology classes participated in this study each composed of mainly 5th year students, age 16/17 (N = 54, Mean = 16.54, SD = 0.50, 44.44% Male), and a very small proportion of 6th year students, age 17/18 (N = 5, Mean = 17.2, SD = 0.45, 100% Male) to give a total of 59 students that participated in the study (Overall Mean = 16.60, Overall SD = 0.53, 49.15% Male). The classes were of mixed ability and had different teachers.

Permission was obtained from the head teacher of the school before this research was carried out. Pupils were required to sign a research consent form to allow the data generated to be analysed; however, every pupil in the classes participated in the activity.

Data Sources

A mixed-method research design was used during this study in order to extract a variety of information. The data sources included a pre-test, post-test and a student attitude questionnaire.

Pre and Post Tests

Four questions, analogous to the types found within the simulated peer-assessment activities, were developed to measure the prior knowledge (pre-test) and resultant knowledge (post-test) of the participants before and after the activities were carried out. The questions in the pre-test and the post-test were analogous and so formed four matched pairs which could be used to assess any increase in students' ability due to the completion of the simulated peer-assessment activities. The full test scripts can be found in the supplementary information. In order to ensure content validity of both the pre and post-tests multiple, practising biology educators contributed to the final design. Furthermore, due to the pre and post-tests containing only analogous questions and not the exact same questions, to prevent prior exposure biasing performance on the post-test, an assessment of the correlation of students' performance on each was measured. This was carried out on a sample group of students who did not participate in the *simulated* peer-assessment activity and a McNemar test of the resulting paired data set revealed that students' performance did not differ in between the tests ($p < 0.001$, $N = 24$).

Student Attitude Questionnaire

This is a self-reported questionnaire that consists of three Likert-scale items. Students were asked to respond using a scale of 1 to 5, where 1 = strongly disagree and 5 = strongly agree. The questions posed to the students were as follows:

(1) Did you enjoy this activity?
(2) Would you have preferred to have been marking one of your classmates work, rather than some simulated work?
(3) Would you have been happy to have your work assessed by others in your class?

The student attitude questionnaire was designed to provide a qualitative understanding into the student's attitudes to the three independent questions, as such, a measure of internal consistency using Cronbach's α was not appropriate. The first question was included to assess an element of the usability of the *simulated* peer-assessment activity in the classroom – activities which students do not enjoy are unlikely to be readily adopted by students or teachers. The latter two questions were included to probe the potential benefits of *simulated* peer-assessment over standard peer-assessment as discussed in the literature review.

Research Procedure

The pre-test was carried out by the students at the beginning of the lesson in which they completed the simulated peer-assessment activity. There was not a time limit put on the students as they completed this but it was completed by all within approximately 15 minutes. Students were not made aware of their performance on this test until after the entire research activity had been completed. The students' then participated in the simulated peer-assessment activity for the remainder of the lesson which lasted another 35 minutes, approximately (lessons are 50 minutes each). The students were informed that they did not need to complete all of the example questions and were encouraged to work at their own pace; however they were guided to ensure they had attempted all the different types of questions. During the next lesson the students carried out the student attitude questionnaire and this was followed by the post-test which again was not subjected to a time limit but was completed in approximately 15 minutes. Through luck of timetabling, rather than experimental design, both the pre-test and post-test were administered in the morning which serves to remove a potential confounding variable.

Data Analysis

Students’ responses to the pre and post-tests were scored in a binary fashion as either correct or incorrect. This generated a correct answer rate for each student on each test and a paired t-test was run to examine any overall difference in performance. The pre and post-test combination also made a series of three matched pairs; as such, a McNemar test was employed. The student’s mathematical ability was estimated from their working grade in Mathematics, which was made available by the mathematics department within the school. This was used as a factor in an ANOVA calculation to determine its influence on students’ performance in the pre-test and in their mean gain. Students’ responses to the questionnaire were analysed by plotting histograms.

Results and Discussion

Pre-test and Post-test

The mean correct answer rate for students’ performance on the pre-test was 36.0% (SD = 21.9%, Table 2) which indicates a fairly low performance on this test. The mean correct answer rate for students’ performance on the post-test was 66.9% (SD = 24.7%, Table 2) and the result of the paired t-test indicated that there was a significant improvement of 30.9% on the post-test compared to the pre-test at a 95% significance level (t = 5.96, p-value < 0.001, Table 2). A Cohen’s d value of 1.32, using a pooled standard deviation, indicates this as being a very large effect size.

Table 2: Comparison of the mean correct answer rate between the pre-test and the post-test.

	Pre-test		Post-test			
N	Mean	SD	Mean	SD	*t*-value	*p*-Value
59	36.0%	21.9%	66.9%	24.7%	8.86	< 0.001

To provide further insight, the pre-test and post-test were subjected to a McNemar test to examine the increase in performance on individual questions (Table 3). It can be seen from table 3 that there is very strong evidence to indicate that there has been a significant improvement in performance in the second and fourth questions in the post-test compared to the pre-test (p-values = <0.001). Question 1 shows strong evidence for significant improvement (p-value = 0.002); whereas, for question 3 there is slightly less strong evidence of improvement (p-value = 0.013). These data suggest that the increase in performance is across all questions encountered in the pre and post-tests rather than just localised to improvement in a subset of the questions.

Table 3: McNemar analysis of pre-test and post-test

Question	1	2	3	4
p-Value	0.002	< 0.001	0.013	< 0.001

Each students' overall score in the pre-test was subjected to an ANOVA using their Mathematics working grade as the factor within the analysis. The working grades are scored as A through to D with A being the highest working grade. The results of this analysis are shown in table 4. This ANOVA is of unbalanced design due to the differing group sizes; however, the data set satisfied the underlying assumptions of independence, normality and homogeneity of variance.

Table 4: ANOVA results using mathematics working grade as a factor

Mathematics Working	*N*	Mean Score	SD	F-value	*p*-Value
A	4	81.3%	12.5%	9.343	< 0.001
B	16	37.5%	22.4%		
C	13	31.7%	17.3%		
D	9	27.8%	15.0%		

The p-value for this ANOVA shows a significant difference in the mean score between the different Mathematics working grades. From the mean scores in the table it can be seen that as the students' mathematical ability increases so too does their mean score in the pre-test. Of note is the much larger mean score of those students with a mathematics working grade of A (81.3%, SD = 12.5%) compared to the others. This might serve to indicate that even those students with mathematical working grades of B or C, and hence expected to be proficient with the more advanced mathematics that is encountered in the Scottish Higher mathematics course, such as calculus, are not comfortable with the use of basic mathematical concepts. This author has previously asserted that students in a high school chemistry setting display only an algorithmic 'understanding' of basic mathematical concepts (Scott, 2012).

Each student's overall gain in performance between the pre and post-test was subjected to an ANOVA, again, using their Mathematics working grade as the factor. The results of this analysis are shown in table 5.

Table 5: ANOVA results using mathematics working grade as a factor

Mathematics Working Grade	*N*	Mean Gain	SD	F-value	*p*-Value
A	4	12.5%	14.4%	7.478	< 0.001
B	16	48.4%	21.3%		
C	13	31.7%	26.2%		
D	9	5.6%	16.7%		

The p-value for this ANOVA, $p < 0.001$, shows a significant difference in the mean gain in performance between the different Mathematics working grades. Those students who were of intermediate mathematical ability, i.e. working grades B and C, demonstrated the greatest increase in performance after taking part in the *simulated* peer-assessment activity. For those of working grade B there was a 48.4% increase (SD = 21.3%) and those of working grade C showed a 31.7% increase (SD = 26.2%). Contrariwise, the students with a working grade A and those with a working grade of D demonstrated a much smaller increase in performance. Those students with a working grade of A had a mean gain of 12.5% (SD = 14.4%) and those of working grade D displayed a 5.6% increase in performance (SD = 16.7%). The small mean score gain of the working grade A students can be easily explained as their score on the pre-test was already very high, 81.3% (SD = 12.5%), thus leaving little room for improvement. Participation in this activity is still thought to be useful for this set of mathematical able students as their understanding of how to answer these questions is surely reinforced. The greatest improvement in performance between the pre and post-tests was seen by those students who were of intermediate mathematical ability. This group of students likely possess enough mathematical fluency through which to identify errors presented to them in the *simulated* peer-assessment activity and to engage in ameliorative discussion in order to effect their improvement. A lack of basic mathematical skills is thought to be the reason that those students of working grade D exhibit a negligible increase performance after participating in the activity. Without the prerequisite understanding of simple mathematical techniques, such as the use of ratios or proportional relationships, then they may not be able to initiate or participate in discussion with others in order to find and fix the errors presented to them in the *simulated* peer-assessment activity.

This pattern in improvement across mathematical ability highlights the importance of the existence of a group of students who are unable to access this activity due to a deficiency in mathematical fluency. This suggests that these students could benefit from dedicated instruction in basic mathematical techniques before being embedded within a biology context.

This panel of statistical assessments indicates that *simulated* peer-assessment can have a positive effect on students' performance in numerical problem solving questions in biology.

Student Attitude Questionnaire

The responses from the Likert-scale items were collated and are displayed have been displayed as histograms. Students were asked to respond using a scale of 1 to 5, where 1 = strongly disagree and 5 = strongly agree. Responses to the question "Did you enjoy this activity?" are presented in figure 2. It can be seen that these responses are left-skewed and this demonstrates that the students thought that the activity was enjoyable overall.

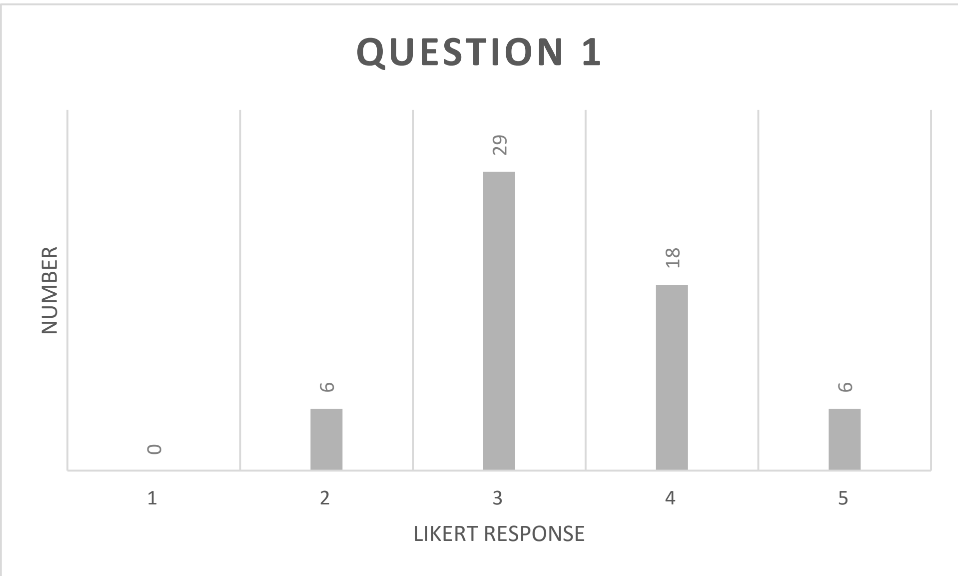


Figure 2: Students' responses to question 1: "Did you enjoy this activity?"

The students' responses to "Would you have preferred to have been marking one of your classmates work, rather than some simulated work" are presented in figure 3.

The responses to this question are right-skewed, which indicates an overall disagreement with this statement; most of the students would not prefer to be marking one of their peer's work. This indicates that the use of simulated material does not result in dissatisfaction for the students which is a finding of note and would be interesting to find out why the students would not have preferred to be marking a peer's work.

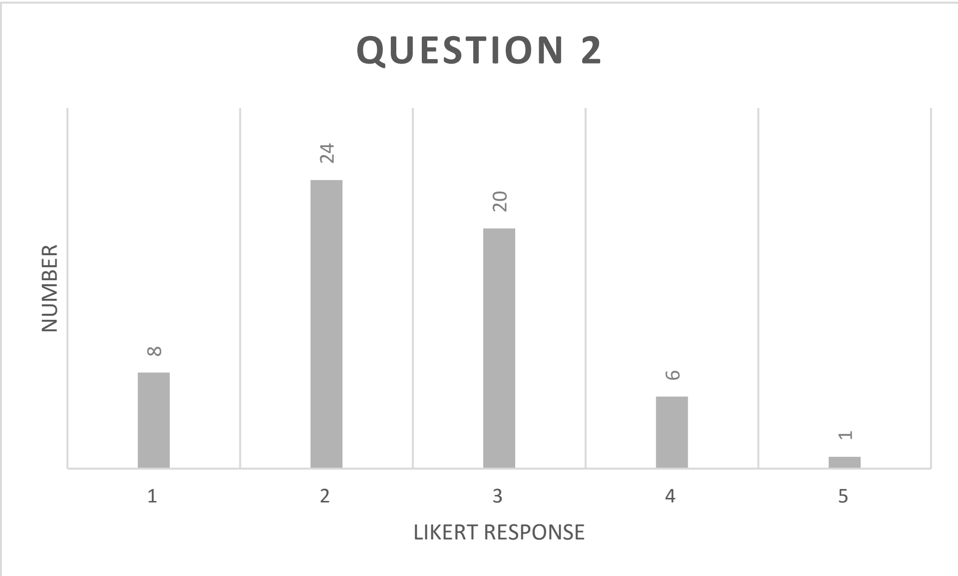


Figure 3: Students' responses to question 2: "Would you have preferred to have been marking one of your classmates work, rather than some simulated work"

The students' responses to "Would you have been happy to have your work assessed by others in your class?" are presented in figure 4. This final question on the student questionnaire shows an overall right-skewed data response with a node appearing at "disagree" on the Likert-scale. This shows that the majority of students would not be happy about their peers assessing their work which is in alignment with the previously asserted disadvantages of peer-assessment.

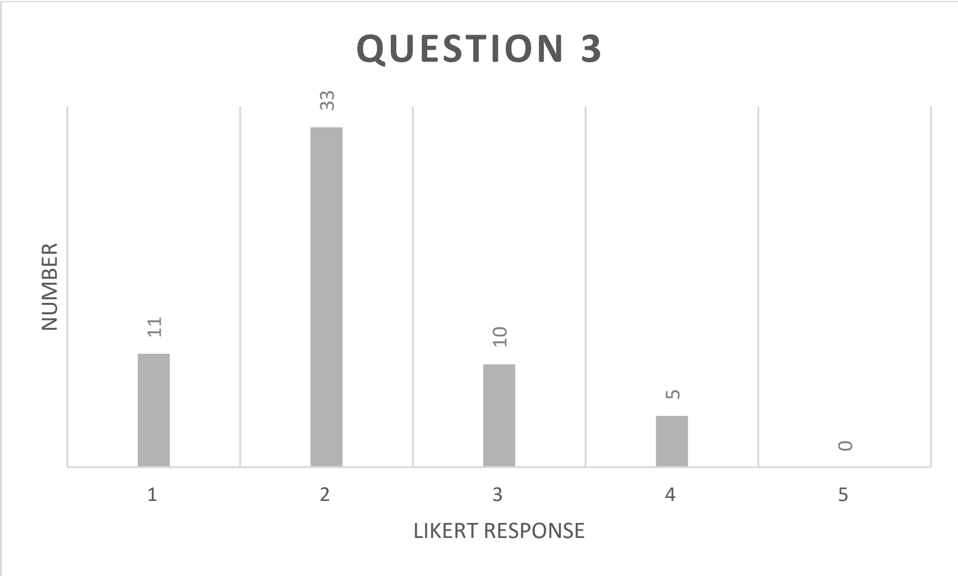


Figure 4: Students' responses to question 3: "Would you have been happy to have your work assessed by others in your class?"

Conclusions

This study has demonstrated that a *simulated* peer-assessment approach can be utilised to promote students' performance in high school biology problem solving calculations. This *simulated* peer-assessment activity has been found to provide the greatest increase in performance to students identified as having intermediate mathematical ability. Furthermore, students of lower mathematical ability would benefit from direct instruction in simple mathematical skills to allow them to benefit fully from participation in this activity.

A student attitude questionnaire was used to probe students' opinion on using a *simulated* design rather than straightforward peer-assessment as it was thought that a simulated approach can be used to circumvent the negative aspects of peer-assessment. The results of the questionnaire have demonstrated that students participating in this study would be less happy to have their work assessed by their peers; as suggested from the literature, this may be due to the previously stated issues of the potential of poor quality feedback or apprehension about their peers viewing their work. Additionally, this study has shown that the replacement of a peer's work with *simulated* work does not result in a negative attitude from the students who took part in this study. These findings would suggest that in the context of high school biology problem solving calculations, *Simulated* peer-assessment can act as a vehicle through which the positive aspects of peer-assessment can be experienced without encountering some of the negative aspects such as poor quality feedback, inconsistent feedback due to social networks within a class or students' apprehension.

The findings of the author's previous study into *simulated* peer-assessment in chemistry context (Scott, 2014) were summarised previously as:

(1) Simulated peer-assessment was able to improve students' performance in chemical calculations.

(2) Students with the greatest improvement were of intermediate mathematical ability.

(3) Students did not display a negative attitude towards the use of 'simulated' rather than 'real' solutions within the activity.

In the present study into simulated peer-assessment in a high school biology problem solving setting, both the statistical evaluation and analysis of the student attitude questionnaire echo the findings of the author's previous study. This study thus serves to expand the evidence for the use of *simulated* peer-assessment as an approach to improve students' performance in numerical areas of science subjects.

Limitations of this Study

There are two main limitations in this study, first the cohort who participated were drawn from only one school and as a result any conclusions made may contain some unwanted bias. The school in question is a well above average in performance on national examinations and so one could speculate that any errors or issues that the students may have in performing the types of questions in this study would be present in schools that perform less well. Similarly, one might also argue that the students within this study, being in an above average performing school, may be more receptive to techniques to improve their performance and thus any increase in performance measured in this study may be over represented. As a consequence of the study taking place in only one school, the sample size was small, although statistically significant results were obtained. Despite these limitations, it is thought that the findings of this study are a valuable contribution to the limited literature on the subject matter.

Implications for the Teaching of Biology

The main drive for this study was to provide research that is both accessible and utilisable to the practicing high school biology teacher. It is thought that this study both highlights the importance of the overlap of mathematics and biology education with regards to numerical problem solving questions and also provides further evidence to support the use of simulated peer-assessment as a means to improve student performance in this area. The activity used in this study is immediately useable to high school biology teachers in Scotland however it is hoped that this strategy can be adapted, as necessary, by practitioners of all levels and locations and incorporated into their teaching practice.

Acknowledgements

The author would like to thank Morag Findlay for her significant assistance in the preparation of this manuscript.

References

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer Interaction and the Learning of Critical Thinking Skills in Further Education Students. Instructional Science, 29(1), 1-32.

Apata, F.S. (2013). Teachers' Experience and Students' Numerical Proficiency in Solving Physics Problems in Secondary Schools. African Research Review. 7(1), 285-297

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. Assessment & Evaluation in Higher Education, 27(5), 427-441.

Bialek, W. and Botstein, D. (2004). Introductory science and mathematics education for 21st-century biologists. Science. 303, 788–790.

Billington, H. L. (1997). Poster presentations and peer assessment: novel forms of evaluation and assessment, Journal of Biological Education, 31(3), 218-220.

Bing, T. J. & Redish, E. F. (2009). Analyzing Problem Solving Using Math in Physics: Epistemological Framing via Warrants," Phys Rev. ST-PER, 5, 020108.

Bing, T.J. (2008). An epistemic framing analysis of upper level physics students' use of mathematics, PhD Thesis, University of Maryland.

Black, P. & William, D. (1998). Assessment and classroom learning. Assessment in Education, 5(1), 7-74.

Bostock, S. J. (2000). Student peer assessment. A workshop at Keele University, May 2000.

Boud, D. (1995). Enhancing Learning through Self Assessment. London: Kogan Page.

Brown, S. & Knight, P. (1994). Assessing Learners in Higher Education, London: Kogan Page.

Bryant, D. A. & Carless, D. R. (2009). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? Educ. Res. Policy Prac., 9(1), 3-15.

Butcher, A. C., Stefani, L. A. J., and Tariq, V. N. (1995). Analysis of Peer-, Self- and Staff-assessment in Group Project Work, Assessment in Education: Principles, Policy & Practice, 2:2, 165-185.

Candy, P. C., Crebert, G., & O'Leary, J. (1994). Developing lifelong learners through undergraduate education. Canberra: Australian Government Publishing Service. National Board of Employment, Education and Training Commissioned Report No. 28.

Carlson, P. A. and Berry, F. C. (2008). Using computer-mediated peer review in an Engineering design course. Transactions on Professional Communication, 51, 264−279.

Chukwuyenum, A.N. and Adeleye, B.A. (2013). Impact of Peer Assessment on Performance in Mathematics among Senior Secondary School Students in Delta State, Nigeria. Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS) 4, 719-725.

DeLorenzo R. (1994). Expanded dimensional analysis, J. Chem. Educ., 71, 789–791.

Denny R. T. (1971). The mathematics skill test (MAST) for chemistry, J. Chem. Educ., 48, 845–846.

Dunbar, K. (1998). Problem solving. In W. Bechtel, & G. Graham (Eds.). A companion to Cognitive Science (pp. 289-298). London: Blackwell.

Eastwood, K.J., Boyle, M.J., Williams, B. & Fairhall, R. (2011). Numeracy skills of nursing students. Nurse Educ. Today, 31, 815–818.

Elwood, J. & Klenowski, V. (2002). Creating communities of shared practice: the challenges of assessment use in learning and teaching. Assessment & Evaluation in Higher Education, 27, 243-256.

Gross, L.J. (2000). Education for a biocomplex future. Science, 28, 807.

Gunersel, A. B., Simpson, N. J., Aufderheide, K. J., and Wang, L. (2008). Effectiveness of Calibrated Peer Review for improving writing and critical thinking skills in biology undergraduate students. Journal of the Scholarship of Teaching and Learning, 8(2), 25-37.

Hammonds,B. (2013). The Effects of Peer and Self-Assessment in a Middle School Mathematics Classroom, Masters Thesis, Ohio University.

Harlen, W. (2007). Assessment of learning. London: Sage.

Heller, P. & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: designing problems and structuring groups. Am J Phys 60, 637–644.

Heller, J.I. & Reif, F. (1984). Prescribing effective human problemsolving processes: problem description in physics. Cogn Instr 1, 177–216.

Hourighan, M. and O'Donoghue, J. (2006). Mathematical Under-Preparedness: The Influence of the Pre-Tertiary Mathematics Experience on Students' Ability to Make a Successful Transition to Tertiary Level Mathematics Courses in Ireland. International Journal of Mathematical Education in Science and Technology, 38 (4), pp. 461-476.

Hsu, L., Brewe, E., Foster, T. M., & Harper, K.A. (2004). Resource letter RPS-1: research in problem solving. Am J Phys 72, 1147–1156.

Karaca, E. (2009). An evaluation of teacher trainees' opinions of the peer assessment in terms of some variables. World Applied Sciences Journal, 6(1), 123-128.

Köhler, W. (1925). The mentality of apes. New York: Harcourt Brace & World. Kroeger, S. (2006). Using Peer-Assisted Learning Strategies to Increase Response to Intervention in Inclusive Middle Math Settings. Teaching Exceptional Children, 38(5), 6-13.

Larkin, J., McDermott, J., Simon, D. P. & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335–1342.

Lazonby, J., Morris, J. E. and Waddington D. J. (1982). The muddlesome mole, Educ. Chem., 19, 109–111.

Liang, J. C., and Tsai, C. -C. (2010). Learning through science writing via online peer assessment in a college biology course. Internet and Higher Education, 13, 242–247.

Lesh, R., & Zawojewski, J. (2007). Problem-solving and modeling. In F. Lester (Ed.), Second handbook of research on mathematics teaching and learning (pp. 763-804). Reston, VA: NCTM.

Lester, F. K., & Kehle, P. E. (2003). From problem-solving to modeling: The evolution of thinking about research on complex mathematical activity. In R. Lesh, & H. Doerr, (Eds.), Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching (pp. 501-518). Mahwah, NJ: Lawrence Erlbaum Associates.

Ljungman, A. G., & Silen, C. (2008). Examination Involving Students as Peer Examiners. Assessment & Evaluation in Higher Education, 33(3), 289-300.

Llamas, A., Vila, F., Sanz, A. (2012). Mathematical skills in undergraduate students. A ten-year survey of a plant physiology course. Biosci Educ., 19.

Lythcott J. (1990). Problem solving and requisite knowledge of chemistry, J. Chem. Educ., 67, 248–252.

Matsuno, S. (2009). Self-, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms. Language Testing, 26(1), 75-100.

McComas, W. (2007). Biology Education under the Microscope: Examining the History and Current State of Biology Education. The Science Teacher, 74(7), 28-33.

McDonald, B. and Boud, D. (2003). The Impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. Assessment in Education, 10(2), 209-220.

National Academy of Science (2011). The Current Status and Future Direction of Biology Education Research, Washington, DC: National Academies of Science Press.

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Newell, A. and Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

O'Donovan, B., Price, M. & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. Teaching in Higher Education, 9(3), 325-335.

O’Shea, J.P. (2003). Mathematical Problem Solving in the Senior Primary School: An Investigation into the Beliefs and Experiences of Class Teachers. MEd thesis, University of Limerick, Ireland.

Pawl, A., Barrantes, A. & Pritchard, D. E. (2009). Modeling applied to problem solving. In: 2009 Physics Education Research Conference, AIP Conference Proceedings, vol. 1179, New York: AIP, 51–54.

Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children. London: Routledgate & Kegan Paul.

Peterson, E. R. & Irving, S. E. (2008). Secondary school students’ conceptions of assessment and feedback, Learning and Instruction, 18(3), 238-250.

Polya, G. 1945. How to Solve It. Princeton, NJ: Princeton University Press.

Polya, G. (1962). Mathematical discovery: On understanding, learning and teaching problem solving: Volume I. New York: John Wiley and Sons, Inc.

Redish, E. F. (2006). Problem solving and the use of math in physics courses. Proceedings of the Conference, World View on Physics Education in 2005: Focusing on Change, Delhi, India, August 21–26.

Resnick & Ford, 1981). The psychology of mathematics instruction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Royal Society of Chemistry, (2009a). www.rsc.org/images/ACMERSCResponse_tcm18-170564.pdf, date accessed 10/11/11.

Royal Society of Chemistry, (2009b). http://www.rsc.org/images/ Bulletin10_tcm18-134605.pdf, date accessed 10/11/11.

Royal Society of Chemistry, (2009c). http://discovermaths.rsc.org, date accessed 10/11/11.

Royal Society of Chemistry, (2012a). Is Maths to Blame? Education in Chemistry, 49(5), 7.

Royal Society of Chemistry, (2012b). Criticism of maths in chemistry grows. Education in Chemistry, 49(4), 3.

Rust, C., Price, M. & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes, Assessment & Evaluation in Higher Education, 28(2), 147-164.

Scott, F. (2012). Is mathematics to blame? An investigation into high school students' difficulty in performing calculations in chemistry, Chem. Educ. Res. Pract., 13, 330–336.

Scott, F. (2014). A simulated peer-assessment approach to improving student performance in chemical calculations. Chem. Educ. Res. Pract., Advance Article.

Schoenfeld, A. H. (1985). Mathematical problem solving. New York: Academic Press.

Science Community Representing Education, (2010) http://www.nationalstemcentre.org.uk/res/documents/page/SCORE%20report.pdf

Stefani, L. (1994). Peer, self and tutor assessment: relative reliabilities. Studies in Higher Education, 19(I), 69-75.

Tariq, V. (2002). A decline in numeracy skills among bioscience under graduates. J. Bio. Ed. 36(2), 76-83.

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities, Review of Educational Research, 68(3), 249-276.

Topping, K. J. (2005). Trends in Peer Learning. Educational Psychology, 25(6), 631-645.

Topping, K. J. (2009). Peer Assessment. Theory Into Practice, 48(1), 20-27.

Trautmann, N. M. (2009). Interactive Learning through Web-Mediated Peer Review of Student Science Reports. Educational Technology Research and Development, 57(5), 685-704.

Tsai, C. -C., and Liang, J. C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. Instructional Science, 37, 293−310.

Tsai, C. -C., Lin, S. S. J., and Yuan, S. M. (2002). Developing science activities through a networked peer assessment system. Computers & Education, 38, 241−252.

Van den Berg, I., Admiraal, W. and Pilot, A. (2006). Peer assessment in university teaching: evaluating seven course designs, Assessment & Evaluation in Higher Education, 31(1), 19-36.

Vazquez, J. (2006). High School Biology Today: What the Committee of Ten Did Not Anticipate. Cell Biology Education, 5, 29-33.

Weisman, R. L. (1981). A mathematics readiness test for prospective chemistry students, J. Chem. Educ., 58, 564–565.

Wen, M. L. & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment, Higher Education, 51(1), 27-44.

West African Examinations Council (2004). Senior school certificate examination. Chief Examiner's report, Lagos Office, Nigeria.

White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course, Asian EFL Journal, 33(1), 1-36.