The Development of Undergraduate Science Students' Scientific Argument Skills in Oral Presentations

Andrea Bugarcic[a,b], Kay Colthorpe[a], Kirsten Zimbardi[a], Hing Wee Su[a], Kelly Jackson[a]

Corresponding author: a.bugarcic@uq.edu.au
[a] School Of Biomedical Sciences, University of Queensland, Brisbane, Australia
[b] Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia

Keywords: scientific argument, oral presentation, rubric, measuring argument quality

International Journal of Innovation in Science and Mathematics Education, *22*(5), 43-60, 2014.

Abstract

The Science Threshold Learning Outcomes (TLOs) developed recently as part of the Learning and Teaching Academic Standards project, reinforce that the ability to develop evidence-based, well-reasoned arguments and to clearly communicate those arguments in a variety of communication modes, are key graduate attributes (Jones, Yates & Kelder, 2011). However, in practice, specific measurement of these skills is limited, particularly in oral presentations. This study describes the initial literature-based development of a rubric for the evaluation of scientific argument in oral presentations (Toulmin, 1958; Sampson, Grooms & Walker, 2009), and the reiterative, data-driven process of refinement of that rubric. The rubric reflects the established framework for the scientific argument, by including criteria for claim, evidence and reasoning, and evaluates these three components across standards that represent the variation within a mid-level undergraduate cohort.

Using this rubric, we evaluated the ability of undergraduate science students to communicate scientific arguments in an oral presentation task in which they presented data acquired from an inquiry-based practical (Bugarcic, Zimbardi, Macaranas & Thorn, 2012). Students demonstrated the ability to make claims, supply evidence and articulate reasoning that linked claims with supporting evidence. However, the standard of these elements was varied, and the structure of students' arguments was not always complete. Using an action-research approach, these initial findings were used to develop student guidelines and alter the curriculum in a subsequent iteration of the course. This intervention resulted in students presenting more complete and higher-quality arguments. Overall, this study reports on the development of the rubric and describes the design and impact of an evidence-driven teaching intervention that enhances students' scientific argument development in oral presentations.

Introduction

The development of advanced critical thinking skills is essential for the current scientific workforce (Abi-El-Mona & Abd-El-Khalick, 2011; Erduran, Simon & Osborne, 2004; Lyons, 2006), driving the requirement for the development of scientific argument as a 'skill' in science graduates. Internationally (Billington, Gilmore, Munro, Parkin & Wainwright, 2007), and in Australia (Jones & Yates, 2011; Krause, Simon & Scott, 2012), higher education quality agencies have developed minimum standards of competency a student should attain at the completion of their degree. In Australia, these threshold learning outcomes (TLOs) state that undergraduate science students should be capable of interpreting and drawing conclusions from scientific data and be able to communicate scientific results, information or arguments in variety of modes (Jones & Yates, 2011). The combination of these learning outcomes reinforces that both the ability to develop evidence-based, reasoned arguments, and to communicate those arguments, are key attributes of science graduates.

Although science graduates must have advanced skills in developing and communicating scientific arguments, many science students have never been explicitly taught how to develop an effective scientific argument before enrolling in university (National Research Council, 1997; Ryder, Leach, & Driver, 1999). Therefore, during their undergraduate degree, as they develop from novice to expert scientists, students must have opportunities to develop skills in creating and communicating scientific arguments, to apply these skills to novel situations and in a variety of contexts, and to effectively utilise scientific arguments in different forms of communication (Thomas & Sandhya, 2007; Meers, Demers & Savarese, 2003). Most importantly, to substantiate the claims that students have developed advanced skills in scientific reasoning by the time they graduate, it is necessary to have tools to measure the quality of scientific arguments, and these appear to be lacking from present literature.

The term "scientific argument" is extensively used in the educational community, but the terminology used to define the elements of scientific arguments has undergone several changes since the first theoretical framework proposed by Toulmin in the 1950's (Toulmin, 1958). Based on extensive literature (Dunbar, 1993; Sampson, Grooms & Walker, 2009; Roberts & Gott, 2010; Sampson & Walker, 2012), we frame scientific argument as the overarching unit that consists of a claim, evidence and reasoning. In this study, claim is defined as 'a statement that requires support from extraneous sources (i.e., evidence)', evidence is 'the measurable outcome(s) used to support the claim', and reasoning is 'the explicit link between the claim and the evidence that describes how the evidence is relevant to the claim, and why it makes the claim believable'.

There are established methods for analysing the quality of scientific arguments in written reports that have been used in several undergraduate settings. Research using these methods indicates that high school and undergraduate students appear competent in the production of claims in written work, but are often fail to substantiate these claims with evidence (Sandoval & Millwood, 2005; Jiménez-Aleixandre, Rodriguez & Duschl, 2000). Further, Ruiz-Primo and collegues (Ruiz-Primo & Furtak, 2006) suggest that, although it is rare for students to refer to evidence in their written work, when they do, the evidence is often of a high standard. Kelly and Bazerman (2003) and Sampson and Walker (2012) suggest that as a result of the absence of evidence, reasoning is also rarely present in students' written scientific arguments and, when present, is often of low standard.

Oral presentations have become an increasingly important communication tool for scientists (e.g., LaBlanca, 2011). As the publication lag in top journals increases, and the speed of scientific discovery accelerates, the advancement of scientific knowledge becomes increasingly reliant on conference presentations to disseminate novel research and foster productive discussion (Toft, 1998). More broadly, the current increasing need for scientists to communicate with expert and non-expert scientific as well as non-scientific audiences (Jones & Yates, 2011), increased online availability of digital media (e.g., podcasts) also increases the need for our science graduates to develop skills in oral communication (TLO 4; Jones & Yates, 2011) to contribute effectively to a scientifically literate society. The literature concerned with teaching students oral communication skills for scientific presentations focuses largely on the physical process of presentation. This includes developing presentation technicalities (e.g., improving components of the presentation, such as introduction and methodology sections) and refining the ways students present (e.g., use of loud and clear voice) (LaBlanca, 2011; Chan, 2011). Only a small number of studies have attempted to develop, and then measure, the quality of scientific argument in oral presentations, and these have only used very simple metrics (for example, Kerr and colleagues use "Convinces me of his/her argument" on a 1 (NO) to 5 (YES) scale) (Haworth & Garrill, 2003; Kerr, Murray, Moore & Nonhebel, 2000). Therefore, the current literature appears devoid of measurement tools suitable for analysing all of the elements necessary to form effective, complex scientific arguments in oral presentations. In our opinion, if we want to develop this skill in our students we need to (i) provide suitable opportunities for this development, (ii) define its elements, (iii) develop clear guidelines for our students and (iv) objective assessment and measurement tools appropriate for oral presentations.

We report here a two-year study during which we used a well-established action research approach which enables educators to use existing educational theoretical and practical frameworks to identify, address and evaluate a problem in their own teaching settings (e.g., McNiff, & Whitehead, 2006a; McNiff, & Whitehead, 2006b; Reason & Bradbury, 2006). Using this action research loop of 'observe – act – evaluate – reflect', in the present study, we identified key problem areas in students' oral scientific arguments, developed a method to measure the quality of scientific argument in oral presentations of undergraduate science students, developed and implemented an intervention to address these key problems, and measured the impact of this intervention on the quality of students' oral scientific arguments. Specifically, we used transcripts of video recordings of undergraduate science students' oral presentations to adapt published rubrics that evaluated the quality of written scientific arguments (Sampson & Walker, 2012; Brown, Nagashima, Fu, Timms & Wilson, 2010), which used a 1-5 standard scale for each of the argument elements: claim, evidence and reasoning. The rubric was then used to compare the quality of the scientific arguments provided by two cohorts of students enrolled in Bachelor of Biomedical Science program, before and after the implementation of the teaching intervention targeted at improving students' construction and oral communication of scientific arguments. The rubric revealed a clear increase in quality of arguments in oral presentation, specifically in claim and reasoning elements as a result of the teaching intervention. Furthermore, the intervention helped students to more clearly articulate what constitutes a good argument and the relationship between the argument elements.

Methodology

Ethical Clearance

Ethical clearance was obtained from the UQ Behavioural and Social Sciences Ethical Review Committee prior to the commencement of this study (clearance no. 2012000472), and participating students provided informed consent for the use of their oral presentation data to be included in this study.

Student cohort details

Oral presentations analysed in this study were provided by undergraduate science students enrolled in *Integrative Cell and Tissue Biology* course (BIOM2013), a compulsory second year course for the Bachelor of Biomedical Science program at the University of Queensland. BIOM2013 is a one-semester (13 week) course, which consists of three 1-hour lectures and one 3-hour practical class each week. As part of the successful completion of the course, it was compulsory for students to participate and complete a practical component (40% of overall course grade) structured around the investigation of the effects of different pharmacological agents on macropinocytosis (Bugarcic, Zimbardi, Macaranas & Thorn, 2012).

Data collection

The assessment for the first part of the practical component required the students to provide a 10-minute group oral presentation explaining their experiments and discussing their findings. These presentations were video recorded and recordings from students who consented to being part of the study were transcribed. Presentations from 2012 and 2013 student cohorts were included in the analysis; eight student groups (3-4 students per group) from 2012 and a further 5 groups (3-4 students per group) in 2013. All other aspects in the course, including assessment, did not change from 2012 to 2013.

Data analysis and representation

Data analysis: Transcripts of the presentations were analysed as described in the Rubric development (Results section) below, to identify statements that represented the three elements of scientific arguments: claim, evidence and reasoning statements. These statements were extracted from the transcripts and aligned in tables to determine the number of complete arguments (containing all three elements) used in the oral presentations of each group. This analysis also enabled extraction and count of the incomplete arguments each group used.

Data representation – Heat map: The quality of the elements in each of the arguments presented by consenting groups from both cohorts (2012 and 2013), extracted and analysed as described above, were evaluated using the rubric developed in this study (Table 1). The quality was scored on a scale of 1-5, with each score also assigned a colour on a red-green scale (1= red, 2 = orange, 3 = yellow, 4 = lime, and 5 = green). These colours were then applied to the tables of arguments that aligned the claim, evidence and reasoning components for each argument, to provide a visual representation of both argument completeness and quality.

Data representation – Bar graph: Colour-coded bar graphs were constructed to visualise the position of argument elements within the context of each oral presentation transcript. Specifically each argument element was again colour coded (claim - orange, evidence – purple, reasoning - blue) and the position of the elements within the total presentation time (denoted as number of words) was represented as a bar graph. Words and statements that were not considered to be elements of arguments are represented as black. Therefore, the colours assigned indicate the length/number of words used for each element within a single transcript, and the bar illustrates the order of elements within the entire transcript.

Meta-learning task analysis

BIOM2013 students in both 2012 and 2013 cohorts were required to complete a meta-learning task following the completion of their oral presentations. The meta-learning task, a survey-type task which asks students to reflect on their content and skills learning and understanding (Biggs, 1985; Hattie, 2009), consisted of five self-reflection questions and completion of this task contributed to 1% of the overall course grade. The five open-ended questions targeted the students' perceptions of their learning in the practical module in the course, with the objective of encouraging students to reflect on their performance and learning within the course. Among the five questions, one question asked the students to comment on their learning gains in oral communication: "Describe one thing you have learned about scientific communication (in oral or written formats) during the practical module and associated assessment". Student responses to this question were analysed using inductive thematic analysis (Hatch, 2002), a qualitative data methodology that identifies and categorises common themes which surface from the data.

Statistical analysis

All quantitative data is expressed as mean ± standard error of the mean (SEM). Multiple Mann-Whitney-Wilcoxon tests with Bonferroni corrections were used to determine whether there were any statistically significant differences between the 2012 and 2013 for the number of complete arguments, and for scores for claim, evidence and reasoning. A student's t-test was used to determine whether there were any statistically significant differences between the 2012 and 2013 cohorts in the proportion of oral presentation words used to provide arguments vs non-argument statements. For each statistical test, an overall p-value of 0.05 or lower was taken to be statistically significant.

Results

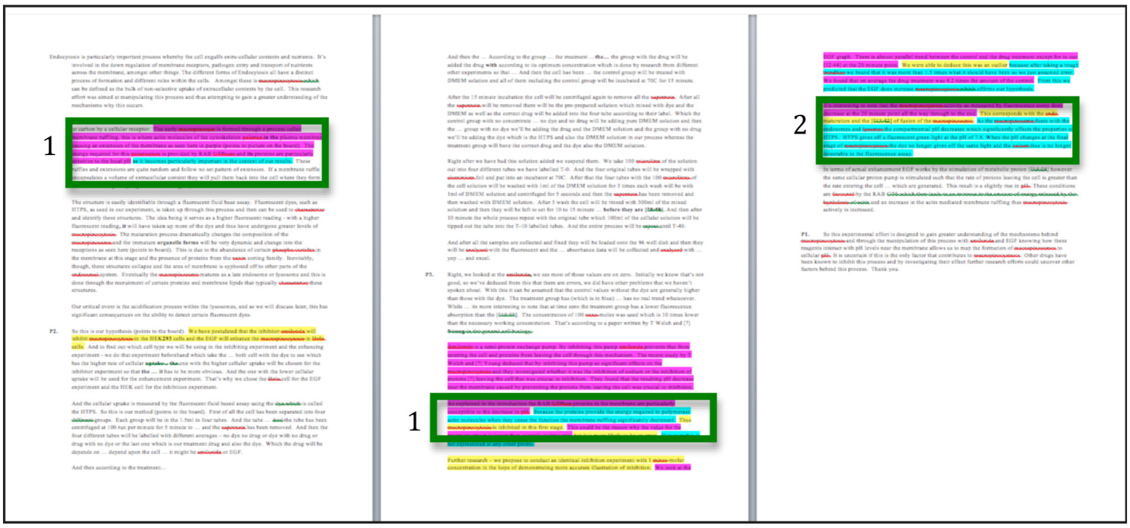
Rubric development using reiterative qualitative analysis

The rubric for the evaluation of the quality of each of the three components of scientific arguments in oral presentations was initially synthesised from the published rubrics of Sampson and Walker (2012) and Brown, Nagashima, Fu, Timms, and Wilson (2010). These publications focused on the evaluation of scientific reasoning within written reports of high school and undergraduate science students, in the context of the three argument elements: claim, evidence and reasoning, using a framework for argumentation initially developed by Toulmin (1958), and adapted by Dunbar (1993) for scientific contexts. To enable the measurement of argument quality in oral presentations, the initial adaptation and amalgamation of the rubrics included (i) refinement so that the rubric developed contained only the key elements of interest (claim, evidence and reasoning), as the rubrics had been used for a broader purpose in their original forms, and (ii) alteration of the key descriptors to better account for the features present in the oral transcripts which differentiated the standards for each of the elements.

This produced the first iteration of the rubric, which when used, revealed the frequency of each argument element used by the 2012 cohort, as well as the qualities of these statements (data not shown). Although statements representing all elements of argument were identified in the 2012 transcripts, there was a distinct lack of evidence statements. This 'missing' evidence suggested that either the students made multiple claims and reasoning statements for each evidence statement, or the analysis had failed to identify all instances of evidence. To elucidate this further, we re-analysed the data, aligning and tabulating each argument found within the transcript (Figure 1A), so that each claim element was linked to its respective evidence and reasoning elements (Figure 1B). This pre-processing of the transcripts for argument analysis had several advantages: (i) it enabled us to effectively score for quality of all elements within a specific linked argument, (ii) it enabled us to find the 'missing' evidence, and (iii) also construct a "heat map" representation of the data obtained (Figure 1B), where each quality level was assigned a specific colour that allowed easier data visualisation and a numerical score for qualitative data analysis. This analysis also highlighted a flaw in the descriptors for the evidence criteria in the earliest iteration of the rubric, and allowed us to develop descriptors for reasoning which critiqued the appropriateness of the links between claims and evidence fundamental to the definition of reasoning in the context of scientific argument frameworks (Toulmin 1958, Dunbar 1993).

Overall, several reiterative cycles of deductive and inductive analysis were used to adapt the rubric to the context of oral presentations, to ensure it accounted for the entire range of qualities present in our transcript data and that it provided adequate differentiation between each of the standards, and to confirm that it was transparent enough to be used consistently by multiple coders with minimal training. Once developed, this rubric (Table 1) was used to analyse argument quality in both 2012 and 2013 cohorts.

A



B

	Claim	Evidence	Reasoning
1	Thus macropinocytosis is inhibited in this first stage.	This could be the reason why the value for the treatment group is lower than controls at time zero The early macropinocyte is formed through a process called membrane ruffling, this is where actin molecules of the cytoskeleton polarise in the plasma membrane causing an extension of the membrane as seen here in purple (points to picture on the board). The energy required for this polarisation is provided by RAB GTPases and its proteins are particularly sensitive to the local pH As explained in the introduction the RAB GTPase proteins in the membrane are particularly susceptible to the decrease in pH.	so it becomes particularly important in the context of our results. Because the proteins provide the energy required to polymerase actin molecules when they cease the function the membrane ruffling significantly decreased.
2	This corresponds with the endo maturation and the [?13:42] of fusion of the macropinosome.	It's interesting to note that the macropinocytosis activity as measured by fluorescence assay does decrease at the 20 minute point all the way through to the end.	As the macropinosome fuses with the endosomes and lysomes the compartmental pH decreases which significantly affects the properties of HTPS. HTPS gives off a fluorescent green light at the pH of 7.8. When the pH changes at the final stage of macropinocytosis the dye no longer gives off the same light and the colour thus is no longer detectable in the fluorescence assay.

Figure 1: Analysis of oral presentation transcripts. Sample colour-coded transcript indicating location of each argument element (yellow=claim, purple=evidence and blue=reasoning) (A) with supporting table of coded arguments (B). The supporting table shows the alignment of coded elements from two different arguments within the transcript in (A) labelled 1 and 2.

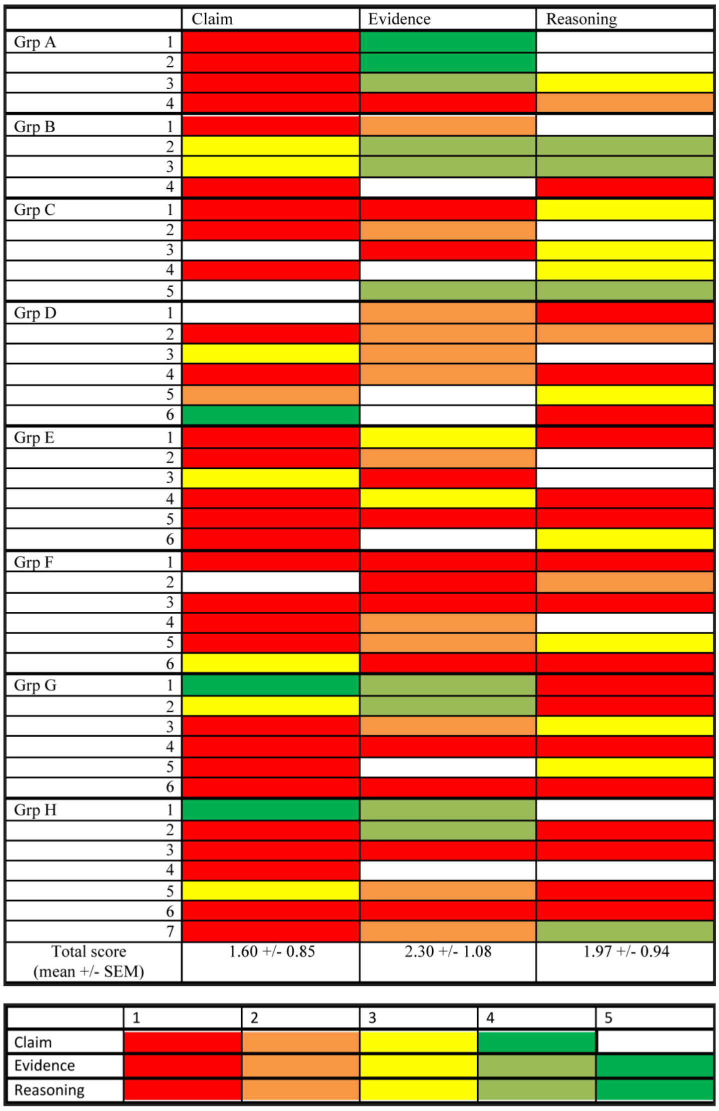
Table 1: Rubric that measures scientific argument quality in oral presentations.

	Standard					
Element	0	1	2	3	4	5
Claim	Claim missing	The author provides a brief claim and lacks major detail of context and variables to address the argument/incorrect detail	The author provides a claim with some detail but it does not include everything that is needed. Major point/important information regarding context left out	The author provides a claim with some detail but it does not include everything that is needed Minor point left out	The author's claim is detailed and includes everything that it should Context stated with necessary units and variables to address the argument	
Evidence	Incorrect	The author did not use data to show trend over time OR a difference between groups (or objects) OR a relationship between variables	The author uses data to show a trend over time OR a difference between group OR a relationship between variables	The author uses data to show a trend over time OR a difference between groups OR a relationship between variables Attempts to support claim AND data is valid but support is incorrect	The author uses data to show a trend over time OR a difference between groups OR a relationship between variables Correctly supports claim AND data is valid but support is incomplete	The author uses data to show a trend over time, a difference between groups (or objects), or a relationship between variables AND included correct units (where appropriate) Completely and correctly supports claims AND data is valid
Reasoning	Reasoning missing	The author provides superficial/broad or incorrect reasoning for the argument	The author provides superficial/broad reasoning that is valid/correct for the argument BUT has missing/incorrect detail/explanation to support argument	The author provides superficial/broad reasoning that is valid/correct for the argument BUT only has some correct detail to support argument	The author provides reasoning that is valid/correct for the argument AND has most of the valid detail required to support the argument	The author provides complete detailed and valid reasoning for the argument

Analysis of arguments in oral presentations of 2012 cohort using the optimised rubric

Using the method for pre-processing the transcripts and the optimised rubric described above, we identified that in 2012, (i) students made, on average, six arguments per oral presentation, (ii) claim was present in most of the arguments but it was of average or low quality, (ii) evidence was present in most arguments and was of slightly higher quality than claims and (iii) reasoning was the element most likely to be missing or not of high overall quality (Figure 2A). Furthermore, analysis of the time students spent on each argument element in the context of a whole oral presentation revealed students spent most of the allocated 10 minutes on the introduction and methodology sections, which were devoid of any argument elements (Figure 2B).

A



Heat map key

B

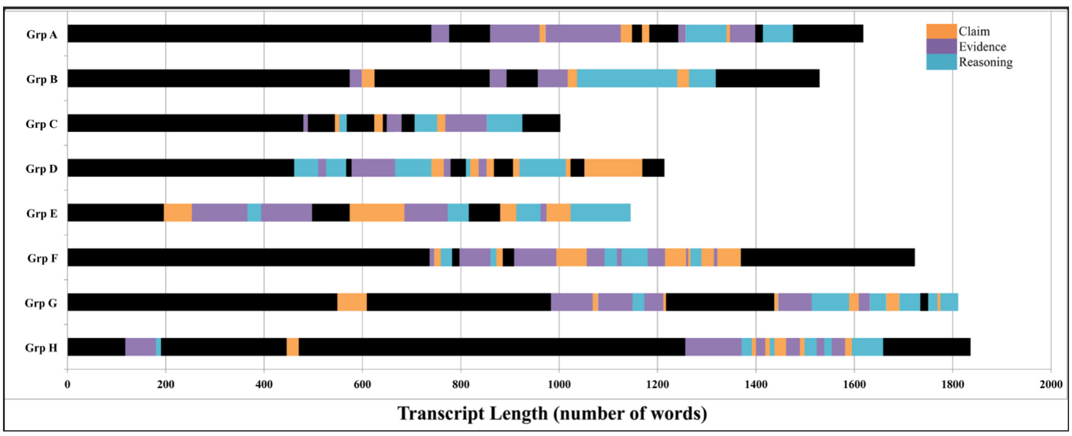


Figure 2: Quality and time analysis of arguments identified in 2012 cohort. (A) Heat map of representing argument quality analysis. (B) Bar graph representing argument time analysis.

Development and implementation of an educational intervention targeted at improving students' oral communication of scientific arguments

Analysis of the 2012 oral presentations revealed that students were providing arguments of relatively low quality, across all three elements of claim, evidence and reasoning. They did not provide evidence or reasoning to support numerous claims, and they spent approximately half of the presentation time describing the underlying concepts and their methods instead of providing scientific arguments related to their experimental approach and findings. To address these issues and explicitly teach students how to effectively construct and communicate high quality arguments within the time allowed, a targeted educational intervention was designed and implemented in 2013. The educational intervention consisted of three components: (i) the introduction and methodology sections were removed from the oral presentation, (ii) the elements of scientific argument were integrated into the criteria sheet used to assess students' oral presentations, and (iii) a workshop was conducted two weeks before the oral presentation. It is important to note that this workshop was conducted in the scheduled class time while students were analysing their data and preparing their presentations for the assessment. The workshop lasted approximately for one hour, where the findings of the 2012 analysis were explained to students, along with how this resulted in changes to the oral assessment task design and marking criteria and standards. Students were also provided with a detailed explanation of the features of high quality arguments, which was supported by an annotated copy of the criteria rubric illustrating what each of the newly-included criteria was trying to develop within the presentations. For example, explanation for the reasoning criterium included: "Has the evidence been lined into a cohesive argument using students data and/or the past literature. Are students only assuming and no evidence exist for their conclusions? If outliers exist – are they mentioned and if so how was this explained?". Furthermore, students were provided with examples of the high and low standards for each criterium that concerned argument elements. For example, to explain the different standards of evidence the students were provided with the following examples and explanations (in italics):

Evidence must include units where appropriate, and evidence should be supported with claims about how that data is valid or reliable.

Low evidence standard:
"So it's just 3.6% compared to 2.7%, again no significance statistically but still interesting to note that there was actually some kind of variance." *Statistically incorrect/unsupported.*

High evidence standard:
"You can see here that our no drug with dye control certainly increased over the 40 minute period while our drug with dye had a very rapid increase which is what we expected because it was an enhancer and then had a very rapidly increase between 20 and 30 minutes which we attributed to be due to the maturation phase [7:49]." *Units are included (minutes), with specific examples of data at 20 and 30 minutes. Data is valid or reliable based on the explicit expectations of author.*

Targeted educational intervention shows a marked increase in quality of scientific argument in oral presentations

Analysis of the transcripts from six oral presentations from the 2013 cohort, following the same method as the final analysis of the transcripts from the 2012 cohort oral presentations, revealed several key improvements in the 2013 presentations (Figure 3) compared with the 2012 presentations (Figure 2). Firstly, although there was no overall significant increase in number of complete arguments ($p=0.79$), there was a statistically significant increase in the quality of the claims and reasoning statements (Figure 3A and Figure 4). Secondly, 2013 students used more of their presentation time to provide scientific arguments (Figure 3B), where the proportion of words spoken by each group that were classified as argument elements was significantly higher (p=0.002) in 2013 (63.75 ± 0.02% of words) compared with the 2012 cohort (38.90 ± 0.05% of words). Lastly, there was no change in the quality of the evidence statements from 2012 to 2013 (Figure 3A and Figure 4, p=0.11), so we hypothesised that in 2013, students focused more on their own experiments and less on the background literature, as evidence sources to support their claims. To test this hypothesis, we categorised the evidence statements into two different types depending on whether the source of the evidence was literature or students' own data. We then investigated whether there were any differences between cohorts in the frequency of these two types of evidence. Interestingly, 2012 students actually used very little evidence from the literature in their arguments, with only 3 out of 41 evidence statements identified as literature-based, while the remaining 38 evidence statements were concerned with students' own experimental evidence. This reliance on experimental evidence as opposed to evidence from the literature was repeated in 2013, with only 5 out of 36 evidence statements being drawn from the literature.

Overall, statistical analysis showed a significant increase in the number of words students used to construct an argument, and quality of two argument elements, claim and reasoning in 2013 compared with 2012, while maintaining the quality of evidence, with no differences identified in the sources of evidence used by the two cohorts, or in the proportion of complete/incomplete arguments.

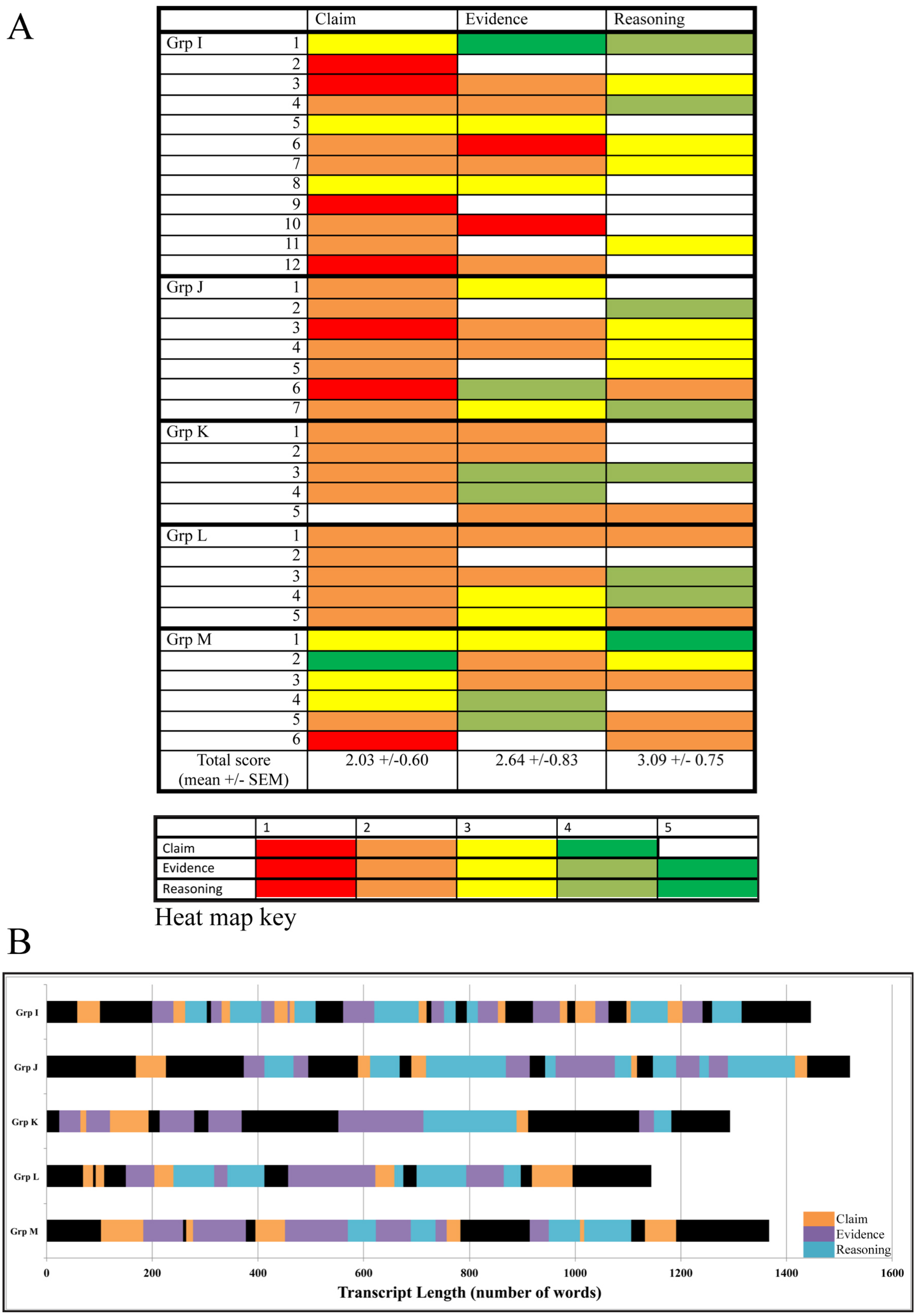


Figure 3: Quality and time analysis of arguments identified in 2013 cohort. (A) Heat map of representing argument quality analysis. (B) Bar graph representing argument time analysis.

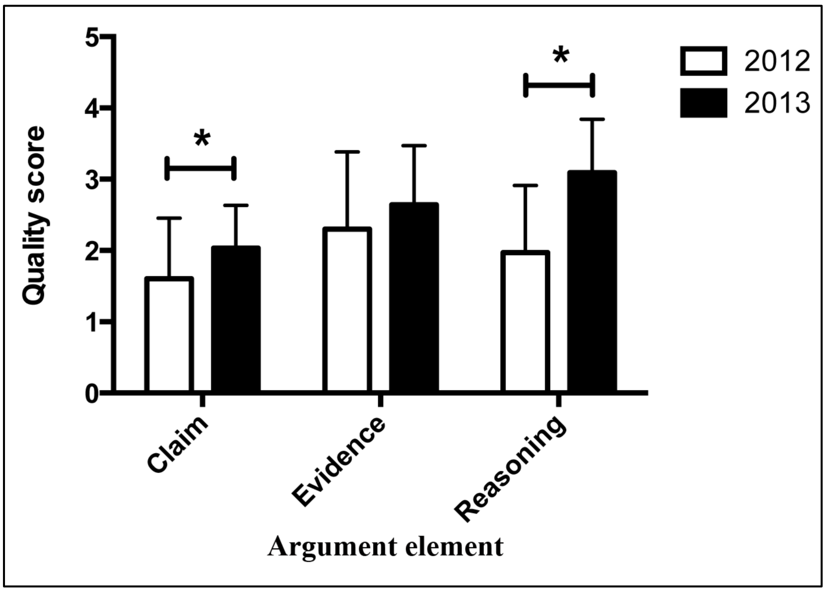


Figure 4: Statistical comparison analysis of argument quality measured in 2013 and 2013 cohorts. Data presented as mean ± SEM score of the quality of each element derived from analysis using the rubric. * indicates 2012 is significantly different to 2013 (p <0.05).

Analysis of student perception of their learning shows an improvement in language use and a greater awareness of argument elements

The students in both cohorts completed a meta-learning task in which they described what they had learnt about oral presentations during the practical module. Overall, analysis of student responses to this open-ended question showed more responses pertaining specifically to using scientific arguments in oral presentation in 2013 (28.8%, 34 out of 118 responses) than in 2012 (15.6%, 15 out of 96 responses). Using inductive thematic analysis, four major themes were identified in responses from both cohorts: "Experimental/human error blame/realisation", "Evidencing/supporting data with facts/references", "Reasoning/explanation of findings/results", and "Evidencing/supporting using data collected". This categorisation revealed that in 2013 a larger number of students indicated learning gains across three of the four themes, than in 2012 (Figure 5). In addition, there were clear differences between the cohorts in the ways that students phrased their descriptions of what they learnt about using scientific arguments in their oral presentation. For example, the representative quotes below clearly show a 2013 student who is aware of the different elements involved in an argument, compared with a 2012 student who describes a similar learning gain using very different language:

> "it is very important to be able to have supporting evidence for all the claims you make" (2013)

> "support your data with logical and scientific facts instead of thinking it happened due to experimental error" (2012)

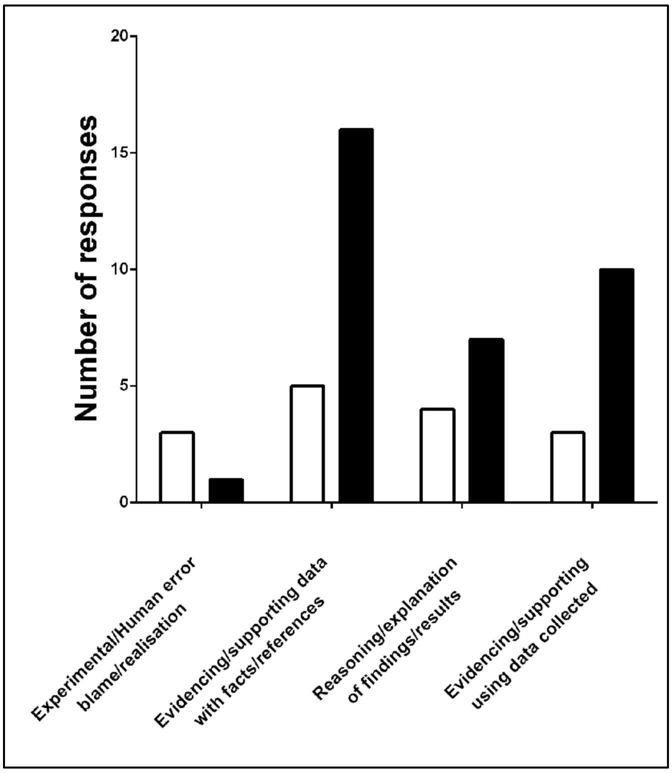


Figure 5. Categorisation and frequency analysis of student responses. Student responses from meta-learning activities in both BIOM2013 2012 (orange) and 2013 (green) were categorised into the identified themes and the number of responses within each theme plotted.

Importantly, the 2013 quote shows recognition that all claims made need to be supported by relevant evidence, while the response taken from the 2012 cohort which does describe an important learning gain in relation to scientific arguments, does not show an understanding of the structure of arguments and specific relationship between elements. Overall, the targeted intervention described here appears to have improved students' articulation of the elements of scientific arguments, and how these elements are related. Thus explicit measurement of these aspects in our study, leading to the explicit teaching of these elements in class, may have enabled students to explicitly recognise and improve this aspect of their oral communication of scientific arguments.

Discussion

This study used an action research approach to (i) observe and identify the difficulties undergraduate biomedical science students had in communicating scientific arguments in oral presentations of their experimental findings from inquiry-style practicals, (ii) develop and implement an intervention targeted at helping students with these specific problems and (iii) evaluate the impact of this teaching intervention on the subsequent cohort's skills in communicating scientific arguments in oral presentations, and in describing their learning about the construction and use of scientific arguments. In doing so, we have not only demonstrated a clear improvement in the quality of students' scientific arguments, but also developed a rubric that measures quality of scientific arguments in oral presentations across standards that describe and differentiate the levels of argument present in 2nd year undergraduate biomedical sciences students. Overall, this study showed that designing educational interventions based on measurable evidence can not only improve student learning outcomes, but can also enable educators to assess differences within and across cohorts.

One of the major outcomes of this study was the construction and data-driven validation of a rubric for measuring the quality of each of the elements of scientific arguments in oral presentations. The current lack of validated tools for evaluating the quality of scientific arguments in oral communication tasks means that it is difficult for educators to demonstrate convincingly that students are developing this essential graduate attribute, and be sure they are facilitating the development of these skills throughout the undergraduate degree. Indeed this important obstacle, limited ability of educators to quickly and accurately evaluate the impact of their curricula, is a recognised problem in advancing teaching and learning (e.g., Elliot, Boin, Irving, Johnson & Galea, 2010). The substantial adjustments we needed to make to published rubrics used to measure the quality of scientific arguments in written work (e.g., Sampson & Walker, 2012) also highlights the importance of validating such tools for different communication modes.

When considering the rubric more generally, it was successful at classifying the scientific argument presented by the 2nd year students who took part in this study. If, for instance, there was a disproportionately high frequency of one of the elements at either extreme end of the rubric standards, then it would be likely that the rubric was not successful in differentiating the standards of the elements of scientific argument in a specific and accurate manner. It is important to note that this disproportionate grouping did not occur, so the rubric was successful in classifying the full continuum of standards in all scientific argument elements within the sample used. As many students enter university as novice scientists, but are expected to graduate with considerable expertise (Jones & Yates, 2011), validation of the rubric on a longer novice-to-expert continuum is an important future study. It is envisaged that when this rubric is used across a broader range of data, for example across the continuity of science students' university degrees, it will enable determination of novice-to-expert range, and thus support educators in diagnosing their incoming students and evidencing the quality of their outgoing graduates.

Identification of arguments in oral presentations, even in written transcripts of the presentations, was not an easy task. While the written work has the ability to be carefully worded, structured and edited multiple times, the pressures of presenting in front of an audience, for an assessment task, has the potential to cause even the most rehearsed presenter to stumble (Joghin, 1999). This makes the construction of the argument in the most logical sequence, claim-evidence-reasoning, difficult. Our analysis of the location of each element of each argument in the transcript has highlighted the extreme messiness of argument construction and cohesiveness in students' oral presentations. For example, while we identified instances of the conventional sequence of claim, followed by evidence, and then reasoning, we also observed the "jumbled" sequences, where students began their arguments with either evidence or reasoning, before continuing on to a second element of the argument that had equal potential of being either of the two remaining elements. Furthermore, we identified the position of the argument elements within the oral presentations may be either disparate or compact. This is represented in Figure 1A, where argument 1 contains "filler" sentences between the elements related to the same argument (i.e., disparate), while argument 2 is an example of a compact sequence where the claim and evidence were linked to the relevant reasoning within the same section of the presentation. The comparison of the "element sequence" and "element positioning" features showed no difference between the

2012 and 2013 cohorts, demonstrating that while the quality of the argument can differ, the way that argument is presented has not been influenced by the particular intervention used in this study. As these important features may themselves influence the quality of the delivered arguments, they may require further targeted interventions and specific guidelines.

As science educators we are in a constant battle between quality of our overall content teaching and time spent on targeted the teaching of specific skills, such as scientific argument in a particular communication medium. While both seem to be of equal importance, in reality, explicit teaching of specific skills is limited in science education, and skill development is often embedded in the curriculum where students are unable to either recognise or comment on it. As discussed above, an important obstacle in this is the limited ability of educators to evaluate the impact of their teaching on student skill development in a quick, easy and longitudinal manner. For example, across several Australian universities, Elliot and colleagues found educators had no objective tools available to them to measure the impact of inquiry-based practical curriculum innovations on student's learning gains in scientific inquiry and problem solving skills (Elliot, Boin, Irving, Johnson & Galea, 2010). Further, our personal extensive experience with Honours students preparing and delivering their first oral presentation (typically a project proposal), serves as a constant reminder that our students are graduating their bachelor degrees with a clear lack of understanding of scientific argument construction and communication. This highlights important gaps both in preparatory education at the undergraduate level, and also the current lack of appropriate tools available to educators to identify, measure and target this learning goal. This experience with Honours students was the primary driver for integrating an oral presentation task in the 2012 cohort's practical assessment. However, the 2012 presentations lacked personal insight and showed superficial use of reasoning, particularly around unexpected results, making it very clear to us that simple integration of assessment task is not enough to drive a specific skill development within it. Therefore, we devised an objective quantitative analysis of a key learning outcome (the construction and communication of sound scientific arguments) that was then used to inform the educational intervention that targeted the specific issues students struggled with in achieving this learning goal.

The first iteration of the educational intervention described herein has focussed on providing students with the same insights we gained from the detailed analysis of the 2012 arguments. The rubric itself was used with students to unpack the important elements of arguments (illustrated by the rubric criteria), and the features that constitute high standard elements and common mistakes illustrated in the descriptors of the lower standards of each element. In the next iteration of the action research cycle, several aspects of argument construction and quality that have not been improved by this intervention will be targeted. Specifically, even though the ratio of words students used for argument construction differed significantly between cohorts, probably due to the removal of introduction and methodology sections, it is important to note that these sections are important for the overall argument delivery within a presentation. For example, the introduction should be developed in a way that it supports and provides a basis for the hypothesis, and highly developed arguments should be present in this section. Furthermore, we were unable to improve the quality of students' use of evidence in their arguments, which may represent the most crucial element of a scientific argument. This, the time spent on argument construction and the cohesiveness of argument elements critiqued above, will form the basis for future interventions and evaluations, using a similar approach and methodology as that described in the present study.

Here we used a mixed-methods approach (Chi, 1997) - a quantitative analysis of qualitative data based on a specifically designed rubric. This approach allows for the specific, objective measurement of each of the three elements essential to scientific arguments: claim, evidence and reasoning. It provides educators with insights into the difficulties students have with construction and communication of scientific arguments. It also provides educators with a way to evidence the impact of teaching interventions targeted at improving students' oral communication of scientific arguments, and a way to provide convincing evidence of the level of skill their students develop within this complex graduate attribute. This approach can therefore be used to inform targeted educational interventions that can, in turn, be evaluated in an objective and measurable way, and be can be used to develop the explicit awareness of the key features of rigorous scientific arguments of both educators and students.

Acknowlegdments

Sincere thanks is extended to the School of Biomedical Science at the University of Queensland for their support of the development of this project, to all tutors for all their help with sucessful implementation of the intervention and to the students that made their presentations available.

References

Abi-El-Mona, I. & Abd-El-Khalick, F. (2011). Perceptions of the nature and "goodness" of argument among college students, science teachers, and scientists. *International Journal of Science Education, 33*(4), 573-605.

Billington, D. B. J., Gilmore, B., Munro, R., Parkin, S., & Wainwright, A. (Eds). (2007). *Biomedical Science,* United Kingdom: Quality Assurance Agency.

Biggs J. (1985). The role of metalearning in study processes. *British Journal of Educational Psychology,* 55(3), 185-212.

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment, 15*(3-4), 142-174.

Bugarcic, A., Zimbardi, K., Macaranas, J., & Thorn, P. (2012). An inquiry-based practical for a large, foundation-level undergraduate laboratory that enhances student understanding of basic cellular concepts and scientific experimental design. *Biochemistry and Molecular Biology Education 40*(3), 174-180.

Chan, V. (2011). Teaching oral communication in undergraduate science: are we doing enough and doing it right? *Journal of Learning Design, 4*(3), 71-79.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(2), 271-315.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In J. E. Davidson & R.J. Sternberg (Eds), *The Nature of Insight*, (pp. 365-395), Cambridge, MA, US: The MIT Press.

Elliott, K., Boin, A. T. B, Irving H., Johnson E., & Galea V. (2010). *Teaching Scientific Inquiry Skills. A Handbook for Bioscience Educators in Australian Universities* Medical Education Unit, University of Melbourne.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915-933.

Haworth, I. S. & Garrill, A. (2003). Assessment of verbal communication in science education. *Biochemistry And Molecular Biology Education, 31*(1), 24–27.

Hatch, J. A. (2002). *Doing Qualitative Research in Education Settings.* Albany: State University of New York Press.

Hattie, J. (2009). *Visible Learning: A Synthesis Of Over 800 Meta-Analyses Relating To Achievement*. New York: Routledge.

Jiménez-Aleixandre M. P., Rodríguez A. B., & Duschl R. A. (2000). "Doing the lesson" or "doing science": argument in high school genetics. *Science Education, 84*(6), 757.

Jones, S. & Yates, B. (2011). *Learning and Teaching Academic Standards (LTAS) Project: Final Report for the Second-Intake Discipline Groups of: Architecture, Building and Construction, Education and Science.* Learning and Teaching Academic Standards (LTAS) Project. Sydney, NSW, Australia: Australian Learning and Teaching Council.

Jones, S., Yates, B., & Kelder, J. (2011). *Science Learning And Teaching Academic Standards Statement* Sydney, NSW, Australia: Australian Learning and Teaching Council.

Joughin, G. (1999). Dimensions of oral assessment and students' approaches to learning. In S. Brown & A. Glasner (Eds.) *Assessment Matters,* (pp 146-156). Buckingham: The Society for Research into Higher Education & Open University Press.

Kelly, G. J. & Bazerman, C. (2003). How students argue scientific claims: A rhetorical-semantic analysis. *Applied Linguistics,* 24 (1), 28-30.

Kerr, W. J., Murray, R. E. G., Moore, B. D., & Nonhebel, D. C. (2000), An integrated communications skills package for undergraduate chemists. *Journal of Chemical Education, 77*(2), 191-194.

Krause, K. -L., Barrie, S., & Scott, G. (2012). *Mapping Learning and Teaching Standards in Australian Higher Education: An Issues and Options Paper.* Teaching and Learning Standards: Issues and options paper.

LaBlanca, F. (2011). The 21st-century oral presentation tool bag. *Science Teacher, 78*(7), 51-55.

Lyons, T., (2006). Different countries, same science classes: Students' experiences of school science in their own words. *International Journal of Science Education, 28*(6), 591-613.

McNiff, J., & Whitehead, J. (2006a*). All you need to know about action research*. London; Thousand Oaks, Calif: Sage.

McNiff, J. & Whitehead, J. (2006b*). Doing action research in education*. London; Thousand Oaks, Calif: Sage.

Meers, M., Demers, N. E., & Savarese, M., (2003). Presenting the scientific process: Introducing philosophy, theory, methods, and ethics. *Journal of College Science Teaching, 33*(1), 34-39.

National Research Council, (1997), *Science Teaching Reconsidered: A Handbook*, The National Academies Press.

Reason, P. & Bradbury H. (Eds.), (2006), *Handbook of Action Research: The Concise Paperback Edition*. London: Sage.

Roberts, R. & Gott, R., (2010), Questioning the evidence for a claim in a socio-scientific issue: an aspect of scientific literacy. *Research in Science & Technological Education, 28*(3), 203-226.

Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: exploring teachers' practices and student learning. *Educational Assessment, 11*(3 & 4), 205–235.

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching, 36*(93), 201-219.

Sampson, V., Grooms, J., & Walker, J. P. (2009). Argument-driven inquiry. *Science Teacher,* 76, 42-47.

Sampson, V. & Walker, J. P. (2012). Argument-driven inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. *International Journal of Science Education, 34*(10), 1443-1485.

Sandoval, W. A. & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Lord, T & Baviskar, S. (2007). Moving students from information recitation to information understanding: Exploiting bloom's taxonomy in creating science questions. *Journal of College Science Teaching, 36*(5), 40-45.

Toft, C. (1998). Oral presentations at scientific meetings. *Herpetologica, 54*(Suppl.), 8.

Toulmin, S. (1958). *The Uses of Argument.* Cambridge, MA: Cambridge University Press.