Using the Writing Template provided by the Science Writing Heuristic (SWH) approach for Quality Arguments

Aeran Choi*
Ewha Womans University

Abstract: This study examined changes in the quality of written arguments produced by freshman students in general chemistry laboratory classes using the SWH approach over a semester; difference in the quality of written argument between the original writing template (year I) and the extended writing template (year II); and any difference between Total Argument and Holistic Argument scores. 140 writing samples from 14 students on the year I and 228 samples from 19 students on the year II were collected. Results indicated that despite fluctuations, the students were producing stronger argument by the end of semester compared to the beginning of the semester. Original SWH template group received significantly higher argument scores than extended SWH template group. For the most of year I laboratory investigations, there was no significant difference in the quality of argument between Total Argument and Holistic Argument scores. An implication of this study would be to provide opportunities for students to practice constructing arguments using the original SWH writing template including questions, claims, evidence, and reflection.

Key words: Written Argument, Inquiry-Based General Chemistry

Introduction

The typical use of writing strategy in general chemistry laboratory class was laboratory notebook. In a traditional chemistry laboratory investigation, instructors provide students with an outline of how to keep a laboratory notebook. The sections of this notebook are typically: title, purpose, procedure, data, analysis of results, discussion, conclusion, and post-lab questions (Bunting, 1999; Ferzli, Carter, & Wiebe, 2005). What are goals for students to write in science laboratory investigations? Several studies regarding the use of writing strategy in science implemented research paper as a way of teaching writing skills (Feldberg, 2007; Firooznia, 2006; Kolikant, Gatchell, Hirsch, & Linsenmeier, 2006; Kroen, 2004; Tilstra, 2001); designed writing assignment questions which stimulate student critical thinking and request students to address explanations (VanOrden, 1990); specially designed writing activity such as describing detailed instruction for reassembling (Reynolds & Vogel, 2007). The idea of these studies including simple laboratory notebooks, research paper, or writing assignment is that writing strategy can promote student learning in science (Bazerman, 1988; Connally, 1989; Keys, 1999a, 1999b, & 2000).

Importantly, as Osborne (2002, p.208) asserts, "Arguing and writing are core activities for doing science." The perspective that emphasizes argument in teaching and learning science was strongly related with the recognition that language is integral part of doing science (Norris & Philips, 2003; Osborne, 2002). Research studies implied that students should be given opportunities of interpreting and making sense of data to develop written argument in scientific inquiry (Lemke, 1990; Takao & Kelly, 2003). With the recognition of the importance of student argument (Osborne, Erduran, & Simon, 2004), this study attempted to look at written arguments constructed using the SWH approach and to see how the quality of students' written argument changes over a semester.

The written arguments analyzed in this study were produced by students who generated their

*Corresponding author: Aeran Choi (achoi@ewha.ac.kr)
**Received on 23 Octorber 2012, Accepted on 29 November 2012

own questions, collected data in the context of scientific inquiry, and developed claims and evidence using an argument-based inquiry approach, that is, the Science Writing Heuristic (SWH) approach (Hand, 2008; Keys, Hand, Prain, & Collins, 1999). The SWH approach was designed to facilitate language use (both in verbal and written) in scientific inquiry as shown in Table 1. The SWH approach provides both teacher and student templates to stimulate verbal argumentation and written argument embedded in scientific inquiry, which are aligned with recommendations by current inquiry-based reforms in science education (NRC, 1996, 2000, 2012). The SWH template for students as shown in Table 1 is a semi-structured writing form that scaffolds student written argument that includes questions, claims, evidence, and reflection. Students in the SWH approach are engaged in active practice of constructing scientific argument as an integral part of science laboratory activity. Students are encouraged to elicit their beginning questions about a topic, identify patterns in their collected data, construct claims based on interpretation of data, support their claims with evidence, and reflect on their investigations. The teacher SWH template as shown in Table 1 guides teachers to facilitate group and class negotiation while students are engaged in scientific inquiry investigation.

In our study, students' writing samples were collected from two different years using the SWH approach. The structure of the writing template for students for year II(extended SWH template group) was slightly different for year I (original SWH template group). While original SWH template includes sections of questions, claims, evidence, and reflection, extended SWH template group used a SWH template with two additional sections, one entitled "Analysis" and the other "Post-laboratory Questions." These additional sections were included at the request of the instructors from the chemistry department, who felt that these two additional sections would help guide students to the main

Table 1
SWH Templates for Teacher and Student

Teacher Template	Student Template
Pre-Laboratory Activities: Teacher engages students to elicit pre-knowledge and gain understanding of the scientific context into which the laboratory is situated.	Questions: What are my questions?
Participation: Teacher encourages students to engage in an inquiry/laboratory investigation.	Test and Collect Data/Observation: What did I do? What did I see?
Negotiation I: Teacher guides students to think about the meaning of their data through journal writing.	Claims: What can I claim?
Negotiation II: Teacher encourages students to negotiate their understandings of the data with their peers. Students are encouraged to make knowledge claims to state explanations for their data.	Evidence: How do I know? Why am I making these claims?
Negotiation III: Teacher assists students to compare their ideas to a textbook and on-line encyclopedia.	Reading: How do my ideas compare with others?
Negotiation IV: Teacher encourages students to communicate their current understandings of the investigation in a more polished form, i.e., writing a poem, letter or report, or creating a presentation or poster.	
Exploration: Teacher engages students to bring reflection to their understanding of the laboratory concepts.	Reflection: How have my ideas changed?

focus of each laboratory investigation. The guideline of the Analysis section was provided to assist students to interpret, organize, and present their data for each laboratory investigation. The Post-Laboratory Questions were provided to guide students to summarize and incorporate their findings. As students followed the guideline of the Analysis and answered the Post-Laboratory Questions, they were requested to incorporate them into their writings using the SWH template. With respect to this, we were interested in examining if the quality of arguments developed by students with the extended SWH template would be different from one by students with the original SWH template.

The analysis of the written argument quality focused on two forms, that is, the Total Argument score and the Holistic Argument score. Each writing sample was given a *Total Argument score* that was the sum of each score for argument components such as questions, claims, questions-claims relationship, evidence, claims-evidence relationship, and reflection (Choi, Notebaert, Diaz, & Hand, 2010). Choi (2010) explains rationale for the argument components identified from students' writing samples produced using the SWH approach. The Total Argument analytic framework not only looks for argument components, but also addresses the ability of the students to identify each element component (questions, claims, evidence, and reflection) of the argument structure and to record each component in the correct category as laid out by the SWH student template. Using this scoring process, we learned that the evidence appearing in reflection section was not counted as part of the Total Argument score. Osborne and Freyberg (1985) asserted that students hold unstable epistemological beliefs concerning what counts as evidence, what counts as data, or what counts as an explanation. Students do not adequately distinguish between theory, hypothesis, and evidence (Carey & Smith, 1993; Driver *et al.* 1996). In this regard, each writing sample was also given a *Holistic Argument score* that evaluated the quality of argument regardless of how the various argument components were categorized or where they were recorded. In order to obtain a high Total Argument score students needed to organize the presentation of their data, claims, and evidence logically, using the each categories laid out by the SWH framework, with a final score being calculated as the sum of the seven component scores. In contrast, the Holistic Argument score was judged by looking at the overall quality of the argument, regardless of where each element component was found or how it was categorized in writing samples. In this respect, we were interested in examining difference between the Total and the Holistic argument scores to see if students could distinguish and record each argument component in the appropriate section, and understand the argument structure, which was identified in the previous study (Choi, 2010; Choi, Notebaert, Diaz, & Hand, 2010).

Research questions of this study are as follows: Does the quality of arguments developed by students engaged in the SWH approach improve over a semester? Does the quality of arguments developed by students using the original writing template (including questions, claims, evidence, and reflection) differ from one developed by students using the extended writing template (with analysis and post lab questions)? Is there difference between the Total Argument score and the Holistic Argument score?

Methods

Data Collection

Participants in this study were thirty-three freshman students from an inquiry-based general chemistry laboratory course at a large university located in the mid-west. The students were required to enroll in both a lecture course and a related laboratory course for general chemistry. Fourteen students participating in the first year of this study were recruited from a single section of the general chemistry laboratory course. They completed and submitted science writings for each of ten laboratory investigations over the course of one semester. 140 writing samples were collected from fourteen students in year I. In the second year of this study, nineteen students from several different sections of the general chemistry laboratory course volunteered. They completed and submitted science writings for each of the twelve laboratory investigations over the course of one semester. 228 writing samples were collected from nineteen students in year II. In total, the 368 science writing samples from 33 students were collected.

Year I students had the original template provided by the SWH approach which includes questions, claims, evidence, and reflection. Year II students, however, used a writing template with two additional sections: an "analysis" section and a "post-lab questions" section. These additional sections were included at the request of the instructors from the chemistry department, who felt that these two additional sections would help guide students to the main focus of each laboratory investigation. The guideline of the analysis section was provided to assist students to interpret, organize, and present their data for each laboratory investigation. The post-lab questions were provided to guide students to summarize and incorporate their findings. As students followed the guideline of the analysis and answered the post-lab questions, they were requested to incorporate them into their writing using the SWH template such as claims, evidence, and reflection.

Although all thirty-three students were enrolled in the same course, i.e., general chemistry laboratory course, the topics of the laboratory investigations for year I were not the same with ones for year II as they were from two different years. Topics for year I were as follows: acids, bases and the preparation of a salt; the empirical formula of an oxide of copper; acid, bases and their reactions; calorimetry investigations; the heat of formation of magnesium oxide; thermodynamics open inquiry lab; determining the rate law of reaction; the effect of catalysts on the rate of decomposition of H2O2; kinetics open inquiry lab; and gas-phase chemical reactions. Topics for year II were as follows: data collection on properties of soda pop; separation of mixtures; the reactions of several elements; determining the identity of a chemical reactant; investigation of the reaction of zinc and iodine; chemical reactions; investigation of the reactivity of metals; interactions of acids and bases; investigating heat exchange in physical processes; investigating heat exchange in chemical reactions, part 1; investigating heat exchange in chemical reactions, part 2; and alka-seltzer®: an application of gas laws.

Data Analysis

Student writing samples were assessed using the scoring matrix based on an analytical and a holistic framework which was developed to evaluate the arguments generated by students using the Science Writing Heuristic (SWH) approach (Choi 2010; Choi, Notebaert, Diaz, & Hand, 2010). In terms of the analytical framework evaluating the quality of student written arguments, students' fidelity to the argument structure embedded in scientific inquiry using the SWH approach was explored. Seven components were evaluated in analyzing the quality of student written argument in an analytic way; questions, claims, questions-claims relationship, evidence, claims-evidence relationship, use of multiple modal representations, and reflection. The sum of these seven component scores represents the *Total Argument* score. Scoring matrix for each argument component was developed to produce Total Argument score as shown in Appendix A. The analytical framework was designed to evaluate student arguments according to the categories set out by the SWH template. Students, in other words, were expected to be able to distinguish their questions, claims, evidence, and reflection from each other and record each in the appropriate section of the science writing. A second method of evaluating student argument, i.e., the holistic framework was developed to assess the strength and coherence of the argument, regardless of how the various argument components are categorized or where they are recorded. In order to obtain a high Total Argument score students needed to organize the presentation of their data, claims, and evidence logically, using the each categories laid out by the SWH framework, with a final score being calculated as the sum of the seven component scores. In contrast, the Holistic Argument score was judged by looking at the overall quality of the argument, regardless of where each element component was found or how it was categorized in writing samples as shown in Appendix B.

Intra-class correlation (ICC) was used to measure the inter-rater reliability (Shrout & Fleiss, 1979) of scoring using the analytical framework. Two raters performed the scoring of the 130 science writing samples collected. One of the raters was the researcher of this study; the other evaluator holds a doctorate in science education and has two years of teaching experiences at a middle school and a high school. The two evaluators initially scored five of the student writing samples to ensure coherence in how to apply the analytical framework. Whenever there was more than 1 point difference in our scores, a discussion was conducted to resolve the discrepancy. To measure inter-rater reliability, a median intra-class correlation coefficient for the ten writings was generated for each argument component: .916 for questions; .938 for claims; .944 for evidence; .937 for multiple modal representations; .959 for reflection; .910 for questions-claims relationship; .898 for claims-evidence relationship; .972 for Total Argument.

Both in tracking overall change in the students' ability to construct an argument during the course of the semester and in examining any difference in argument scores using original SWH template and extended SWH template, conceptual complexity of each laboratory investigation emerged as a variable that affects the quality of argument that students construct. An instructor of the general chemistry lecture course that was related to the laboratory course rated the conceptual complexity of each laboratory investigation on a scale from 1 to 10. The mean conceptual complexity was 6.25 for the ten labs conducted by original SWH template group, and 5.17 for the twelve labs conducted by extended SWH template group. Using these ratings, each Total Argument score and Holistic Argument score was converted into a weighted score, which accounted for the conceptual complexity of each laboratory investigation (If the Holistic Argument score was 7 and the conceptual complexity is 6, the weighted Holistic Argument score was calculated as 7 times 0.6 and converted into 0.42). The weighted argument scores could then be compared to see whether there is any improvement in the scores over the course of the semester.

In examining any difference between the Total Argument score and the Holistic Argument score, each raw argument score was converted into a percentage score because those two scores are from different scales (If the Holistic Argument score was 6 out of 10, the percentage Holistic Argument score was calculated as 60. If the Total Argument score was 18, which is out of 36, the percentage Total Argument score was 50). The differences between the Total Argument score and the Holistic Argument score were examined using a paired samples t-test analysis for each laboratory investigation for 10 labs for Year 1 and 12 labs for Year 2.

Data analyses were carried out using the Statistical Package for Social Science (SPSS) for Windows, Version 15.0.

Results

Changes in Argument Scores over a Semester

In order to track any overall change in the students' ability to construct an argument during the course of the semester, the argument scores were converted to weighted argument scores to account the difference in conceptual complexity among ten or twelve laboratory investigations. Figure 1 presents the weighted Total Argument scores and the weighted Holistic Argument scores for year I students in each of the ten laboratory investigations. Figure 2 presents the weighted Total Argument scores and the weighted Holistic Argument scores for year II students in each of the twelve laboratory investigations. For year I students, in spite of the fluctuations in the argument scores, there is a tendency for the scores to increase overall. However, the exceptions were the scores for the laboratory investigations 3, 6, 9, and 10. Like year I students, year II students' argument scores tended to improve overall despite some fluctuations in the scores. Figure 2 also shows that students' argument scores for the latter three-quarters of the semester were much better than for the first quarter of the semester.

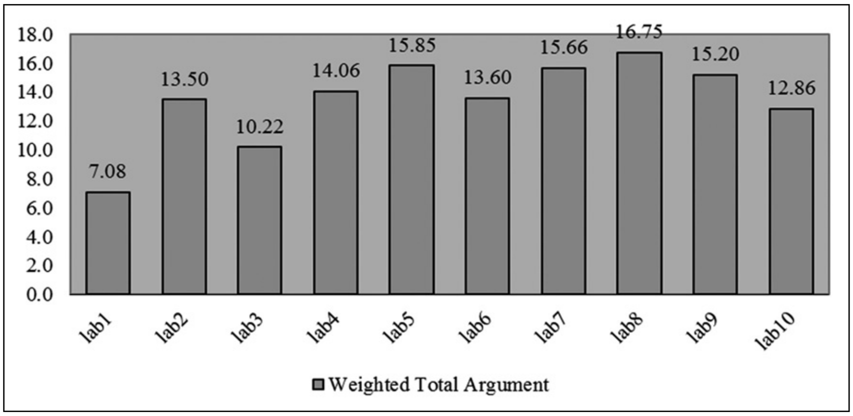

Fig. 1 *The Weighted Total Argument Scores for Year I Students for Each of the Ten Laboratory Investigations*

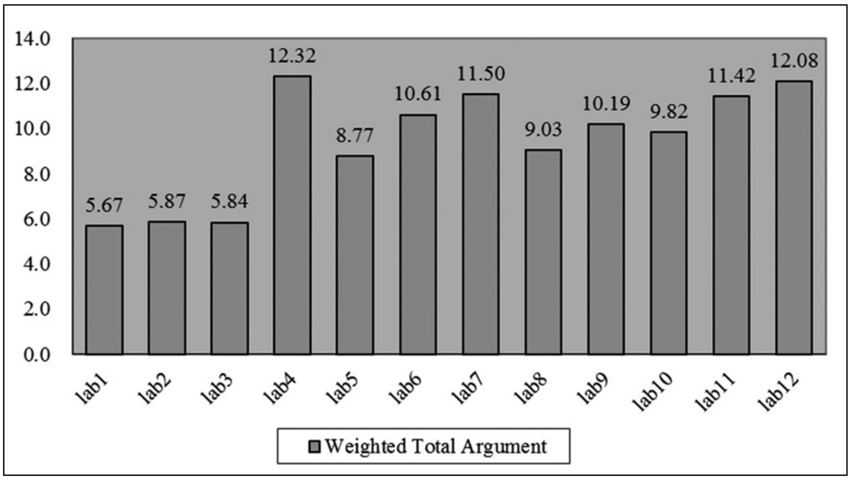

Fig. 2 *The Weighted Total Argument Scores for Year II Students for Each of the Twelve Laboratory Investigations*

Difference in Argument Scores using the Original SWH Template (Year I) and the Extended SWH Template (Year II)

To investigate the difference in argument scores using original SWH template and extended SWH template, an independent-samples t-test was conducted. The raw Total Argument and Holistic Argument scores were converted into weighted Total Argument and Holistic Argument scores which took into consideration the difference in the conceptual complexity of the laboratory investigations between the two years. Once the scores are weighted to account for the different levels of conceptual complexity, the Total Argument scores for original SWH template group are seen to be higher than one for extended SWH template group, as shown in Figure 3. The same is true for the Holistic Argument scores as shown in Figure 4. As shown in Tables 2 and 3, the t-test analysis shows that original SWH

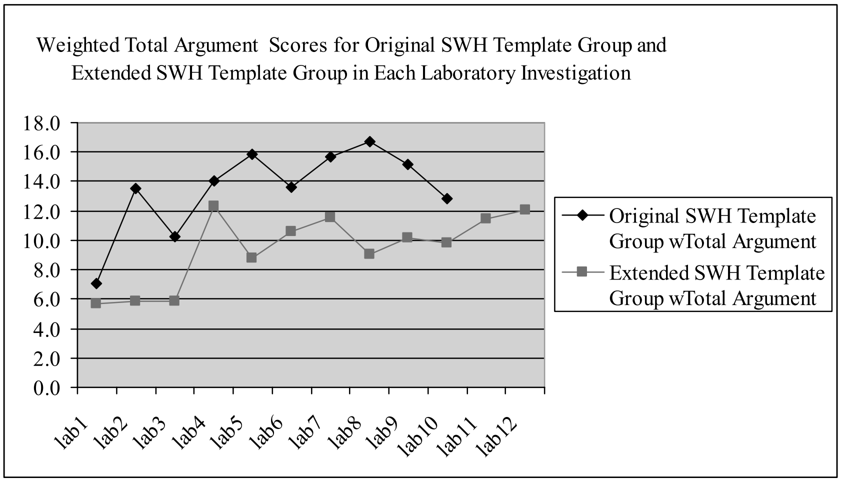

Fig. 3 *The Weighted Total Argument Scores for original SWH template group and extended SWH template group in Each Laboratory Investigation*

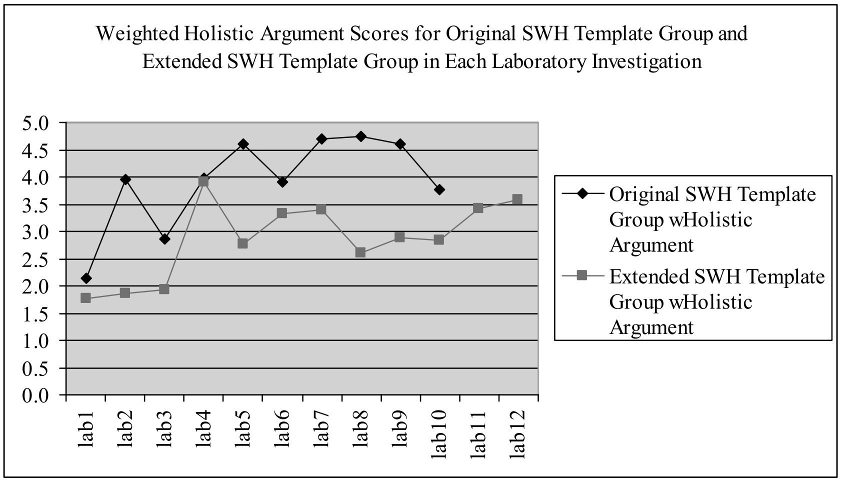

Fig. 4 *The Weighted Holistic Argument Scores for original SWH template group and extended SWH template group in Each Laboratory Investigation*

template group received significantly higher argument scores than extended SWH template group, as measured by both the Total Argument and the Holistic Argument.

Difference between the Total Argument Score and the Holistic Argument Score

For year I students, there were no significant difference between the Total and the Holistic Argument score for the most of the laboratory investigations as shown in Table 4. In two laboratory investigations out of ten, year I students obtained a significantly (p=.05) higher average Holistic Argument score than Total Argument score. For Year II students, for 8 out of 12 laboratory investigations the Holistic Argument scores are significantly higher (p=0.05 level) than the Total Argument scores as shown in Table 5. Higher Holistic Argument score more often occurred in year II than year I.

Discussion

Improvement of Written Arguments over a Semester

Despite fluctuations, the students were producing stronger argument by the end of semester

Table 2

Independent Samples t-test for Original SWH Template Group and Extended SWH Template Group on Weighted Total Argument Scores

Template	N	Mean	SD		Levene's Test for Equality of Variances		t-test for Equality of Means				
					F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
Original	140	134.77	50.95	Equal variances assumed	.005	.942	7.65	366	.00	40.51	5.29
Extended	228	94.25	48.29	Equal variances not assumed			7.55	282.1	.00	40.51	5.36

Table 3

Independent Samples T-test for Original SWH Template Group and Extended SWH Template Group on Weighted Holistic Argument Scores

Template	N	Mean	Std. Deviation		Levene's Test for Equality of Variances		t-test for Equality of Means				
					F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
Original	140	39.23	15.83	Equal variances assumed	.436	.509	6.60	366	.000	10.66	1.61
Extended	228	28.56	14.52	Equal variances not assumed			6.47	274.83	.000	10.66	1.64

Table 4

Paired Samples t-Test for Year I Students

Paired Differences	Mean	Std. Deviation	t	df	Sig. (2-tailed)
pHolArgue1 − pTotalArgue1	4.37	6.31	2.588	13	.023*
pHolArgue2 − pTotalArgue2	3.21	7.56	1.591	13	.136
pHolArgue3 − pTotalArgue3	.40	4.74	.313	13	.759
pHolArgue4 − pTotalArgue4	1.35	7.98	.633	13	.538
pHolArgue5 − pTotalArgue5	2.82	6.32	1.669	13	.119
pHolArgue6 − pTotalArgue6	1.87	6.14	1.136	13	.276
pHolArgue7 − pTotalArgue7	4.21	9.11	1.727	13	.108
pHolArgue8 − pTotalArgue8	1.39	5.10	1.019	13	.327
pHolArgue9 − pTotalArgue9	5.40	7.66	2.635	13	.021*
pHolArgue10 − pTotalArgue10	3.33	6.46	1.931	13	.076

Note: pHolArgue1− pTotalArgue1 equals the difference between the percentage Holistic Argument score and the percentage Total Argument score for the general chemistry laboratory investigation 1. *significant at .05 level

Table 5

Paired Samples t-Test for Year II Students

Paired Differences	Mean	Std. Deviation	t	df	Sig. (2-tailed)
pHolArgue1 − pTotalArgue1	6.46	8.82	3.194	18	.005*
pHolArgue2 − pTotalArgue2	5.53	7.48	3.220	18	.005*
pHolArgue3 − pTotalArgue3	10.12	9.84	4.480	18	.000*
pHolArgue4 − pTotalArgue4	8.25	11.72	3.067	18	.007*
pHolArgue5 − pTotalArgue5	6.58	9.29	3.086	18	.006*
pHolArgue6 − pTotalArgue6	6.14	7.06	3.791	18	.001*
pHolArgue7 − pTotalArgue7	4.01	7.18	2.432	18	.026*
pHolArgue8 − pTotalArgue8	1.96	6.78	1.260	18	.224
pHolArgue9 − pTotalArgue9	1.32	8.48	.676	18	.507
pHolArgue10 − pTotalArgue10	1.90	8.08	1.026	18	.319
pHolArgue11 − pTotalArgue11	3.63	7.55	2.092	18	.051
pHolArgue12 − pTotalArgue12	3.10	4.53	2.980	18	.008*

Note: pHolArgue12− pTotalArgue12 equals the difference between the percentage Holistic Argument score and the percentage Total Argument score for the general chemistry laboratory investigation 12 *significant at .05 level compared to the beginning of the semester. This finding implies for science instructors to continue to provide students with opportunities for constructing arguments in the context of scientific inquiry so that they develop abilities of constructing reasonable written arguments.

One possibility for the exceptions to the tendency of improvement over a semester would be that the students found these laboratory investigations needed more guidance for completing them and constructing high quality of argument. For example, laboratory 6, entitled "thermodynamic open inquiry" and laboratory 9, "kinetic open inquiry" both required students to formulate their own questions and to design their own experiments, which is the highest level of inquiry. Both laboratory investigations could be described as open inquiries, requiring more independence in the investigations, and less guidance than other laboratory investigations. For laboratory 10, "gas phase chemical reactions," many students noted in their reflection section that they found the investigation challenging because the materials were in a gaseous state. For these laboratory investigations, student argument scores were generally lower than the prior one.

The scores for year II generally also tended to improve, with exception of lab 5 and lab 8. For laboratory investigation 5, "investigation of the reaction of zinc and iodine" and laboratory investigation 8, "acids and bases and their reactions," students were required to work through a multi-step mathematical process and to propose the results of these calculations as their claims. Most students simply used a description of this arithmetic process as their evidence for an argument. The quality of argument for these laboratory investigations turned out to be lower than for other laboratory investigation.

Original SWH Template and an Extended format including Post-lab Questions

The argument scores for original SWH template group and extended SWH template group tell us something important about the particular writing framework used. As described in methods section, both original SWH template group and extended SWH template group used the original SWH template as the basis from which they constructed their arguments. However, extended SWH template group was given a SWH template which had two additional sections, one called "analysis" and the other called "post-laboratory questions." Extended SWH template group, who used the SWH template with the additional sections, i.e., "analysis" and "post-lab questions," obtained lower argument scores overall, indicating that they might have more difficulty in constructing arguments. Year I students using the original writing template perhaps focused their attention more precisely and effectively on the key components (questions, claims, evidence, and reflection) necessary to a reasonable argument. The additional sections of extended SWH template were considered to be important by the instructors of the chemistry department who felt that the questions under "analysis" and "post-lab questions" would help guide students to the main focus of the investigation. However, only a few students of extend SWH template group captured the key points from the guideline of the "analysis" and the "post-lab questions" and developed strong evidence or reflection. It was noticed that many students of extended SWH template group paid less attention to the key areas of claims, evidence, and reflection. They concentrated instead on just answering the "analysis" and "post-laboratory questions" sections and ended up with vague and incomplete arguments. Perhaps this was because of the lack of focus on the argument structure such as questions, claims, evidence, questions-claims relationship, and claims-evidence relationship. These additional sections on the SWH template used by extended SWH template group may have distracted them from the fundamental structural and rhetorical features of reasonable scientific writing. This also implies that when students are still grappling with understanding the key elements of an argument they are easily derailed. Extended SWH template group' orientation to only answering "analysis" and "post-lab questions" could be a result of their experience in more typical science classrooms. Students tend to pay more attention to questions or cues that are familiar to them. An implication of this result would be that the original SWH writing template which students are given may affect the overall quality of the argument produced. As a tool for supporting students to construct arguments, questions, claims, evidence, and reflection structure would be good to scaffold the student reasoning and build their sense of what is a reasonable argument.

Total Argument and Holistic Argument Scores

The t-test analysis of the Total Argument scores and the Holistic Argument scores for year I students shows that for most of the ten labs the Total Argument score was statistically the same as the Holistic Argument score. Students have abilities of constructing an argument using the original template provided by the SWH approach by recording each argument component in the appropriate section. An implication of this finding would be for science educators to use written argument structure to help students understand the meaning or purpose of each element component category (such as claims, evidence, and reflection) or how each functions to structure an argument (such as QC relationship and CE relationship).

In two laboratory investigations out of ten, year I students obtained a significantly (p=.05) higher average Holistic Argument score than Total Argument score. Higher score in the Holistic Argument would indicate that although students did not put them in the correct category they were able to make claims or provide evidence. If students present their claims in the evidence section or present their claims and evidence in the reflection section, they receive no points for their Total Argument score, but get credit when their argument was scored holistically. As long as there were claims, evidence, and a connection between those, somewhere in their writing, students could obtain a reasonable Holistic Argument score though a relatively low Total Argument score. When students are engaged in writing activity embedded in an argument-based inquiry approach, they need to be provided with more guidance during the inquiry to help with the write up of an argument with claims and evidence.

The higher Holistic Argument scores than the Total Argument scores occurred in 8 labs out of 12 for Year II students, while in only 2 labs for Year I students. It is worth noting that Year II students in this study used a writing template which has two additional sections: "analysis" and "post-lab questions." While Year II students answered those two traditional sections, they often did not include the appropriate information in sections of claims, evidence, and reflection. By moving to impose written argument components more reflective of traditional laboratory reports, that were not part of the argument based inquiry process, would have generated uncertainty and thus confusion. This implies that students may have difficulty in organizing and clarifying their ideas along the lines of the framework provided by the SWH template.

Constructing a good argument is not a simple task. Thus students need guidance and support that scaffolds the process, and builds their sense of what is a reasonable argument (Bereiter & Scardamalia, 1987). Importantly, the frame - the questions, data, claims, evidence, and reflection as categories in the SWH template would provide essential prompts for constructing a written argument.

This study was limited in two important ways. First limitation of this study was that the research of this study did account for the difference in topics of laboratory investigations with only complexity of the laboratory investigations that was rated by a chemistry professor. There might be other factors that would influence the quality of written argument developed by students across laboratory investigations. Second, the research of this study did not examine any impact of instructional practices that the laboratory course instructors implemented. The quality of implementing the SWH approach would be a factor affecting students' engagement in the SWH approach and the quality of written arguments. Throughout the discussion, the research of this study has suggested an argument-based inquiry approach, for instance, the SWH approach that provides student with opportunities to construct argument in the context of scientific inquiry. Clearly further investigation into the teacher implementation of an argument-based inquiry approach and students' construction of arguments is needed.

References

Bazerman, C. (1988). Shaping written knowledge: *The genre and activity of the experimental article in science*. University of Wisconsin Press, Madison, WI.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bunting, R. K. (1999). Precise writing for a precise science. Journal of Chemical Education, 76 (10), 1407-1408.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, 28, 235-251.

Choi, A. (2010). Argument structure in the Science Writing Heuristic Approach. *Journal of the Korea Association for Research in Science Education*. 30(3), 323-336.

Choi, A., Notebaert, A., Diaz, J., & Hand, B. (2010). Examining arguments generated by Year 5, 7, and 10 students in science classrooms. *Research in Science Education*. 40(2), 149-169.

Connally, P. (1989). Writing and the ecology of learning. In P. Connally & T. Vilardi (Eds.), *Writing to learn mathematics and science* (pp. 1-14). New York: Teachers College Press.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Philadelphia: Open University Press.

Feldberg, R. (2007). Biosocial problems in contemporary America: A course on the use and misuse of scientific knowledge. *Journal of College Science Teaching. July/August*, 35-39.

Ferzli, M., Carter, M., and Wiebe, E. (2005). LabWrite. *Journal of College Science Teaching*. November/December, 31-33.

Firooznia, F. (2006). Giant ants and walking plants: Using science fiction to teach a writing-intensive, lab-based biology class for non-majors. *Journal of College Science Teaching*. March/April, 26-31.

Hand, B. (Ed.). (2008). *Science inquiry, argument and language*. Sense Publishers, AW Rotterdam, The Netherlands.

Keys, C. W. (1999a). Language as an indicator of meaning generation: An analysis of middle school students' written discourse about scientific investigations. *Journal of Research in Science Teaching*, 36(9), 1044-1061.

Keys, C. W. (1999b). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science, *Science Education*, 83, 115-130.

Keys, C. W. (2000). Investigating the thinking processes of eighth grade writers during the composition of a scientific laboratory report. *Journal of Research in Science Teaching*, 37(7), 676-690.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the Science Writing Huerisitic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, 36(10), 1065-1084.

Kolikant, Y. B., Gatchell, D. W., Hirsch, P. L., & Linsenmeier, R. A. (2006). A cognitiveapprenticeship-inspired instructional approach for teaching scientific writing and reading. Journal of College Science Teaching, November/December, 20-25.

Kroen, W. (2004). Modeling the writing process. Journal of College Science Teaching, November/December, 50-53.

Lemke, J. L. (1990). *Talking science: language, learning, and values*. Norwood, NJ: Ablex.

National Research Council (NRC). (1996). *National science education standards*. Washington, D.C.: National Academy Press.

National Research Council (NRC). (2000). *Inquiry and the national science education standards*. Washington, D. C.: National Academy Press.

National Research Council (NRC). (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D. C.: National Academy Press.

Norris, S. P., and Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87, 224-240.

Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education*, 32(2), 203-218.

Osborne, R. & Freyberg, P. (Eds.). (1985). *Learning in science: The implications of children's science*. London:Heinemann.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. Journal of Research in Science Teaching, 41(10), 994-1020.

Reynolds, J., and Vogel, S. (2007). Precisely! A writing exercise for science and engineering classes. *Journal of College Science Teaching. March/April*, 31- 33.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass Correlations: Use in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.

Takao, A. Y. & Kelly, G. J. (2003). Assessment of evidence in university students' scientific writing. *Science & Education*, 12, 341-363.

Tilstra, L. (2001). Using journal articles to teach writing skills for laboratory reports in general chemistry. Journal of Chemical Education, 78(6), 762-764.

VanOrden, N. (1990). Is writing an effective way to learn chemical concepts? Journal of Chemical Education, 67(7), 583-585.

Appendix A. Scoring Matrix for the Total Argument

Scoring Matrix for Questions

point	criteria
1	• Single question • Closed–ended question • Questions are not testable • Unimportant and poor questions • Questions do not capture the essence of the investigation • Questions are insignificant • Questions are of low quality
2	• A few questions • Closed–ended questions • Testable or may be difficult to test • May not be meaningful questions • Questions may not capture the essence of the investigation • Questions may not be significant and adequate • Questions may be of low quality
3	• Multiple questions which are primarily closed–ended questions • If only one, it is open–ended question. • Testable and meaningful questions • Questions may match the essence of the investigation • Questions may be significant and adequate • Questions may be of high quality
4	• Multiple questions which include at least one open–ended question • Testable questions • Questions capture the essence of the investigation • Questions are significant and adequate • Questions are of high quality
5	• Multiple questions which include more than one open–ended question • Testable /scientific questions • Questions capture the essence of the investigation thoroughly • Questions are very significant and adequate • Questions are of very high quality

Scoring Matrix for Claims

point	criteria
1	• Single Claim • Claims are not based on any data or observation • Claim does not capture the essence of the investigation • Claim is insignificant • Claim is invalid and inaccurate • Claim is of low quality

2	• Single or multiple claims • Claims may not appear to have come from their experimental observation/data • Claims may not capture the essence of the investigation • Claims may not be significant and adequate • Claims may not be valid and sound • Claims may be of low quality
3	• Single or multiple claims • Claims may be from their experimental observation/data • Claims may match the essence of the investigation • Claims may be significant and adequate • Claims may be valid and sound • Claims may be of high quality
4	• Multiple claims • Claims are from the interpretation of their experimental observation/data • Claims capture the essence of the investigation • Claims be significant and adequate • Claims be valid, sound, and accurate • Claims be of high quality
5	• Multiple claims • Claims are from and based on the interpretation of their experimental observation/data (Claims about what they found out) • Claims capture the essence of the investigation thoroughly • Claims are very significant and adequate • Claims are very valid, sound, and accurate • Claims are of very high quality

Scoring Matrix for Questions-Claims Relationship

point	criteria
1	• Very weak connection between questions and claims • Claims without any questions or questions without any claims • Questions and claims do not fit at all
2	• Weak Connection • Questions and claims fit loosely • Student develops claims for a few of the generated questions • Claims are uncertain in answering questions
3	• Moderate Connection • Questions and claims fit reasonably • Student develops claims for some of the generated questions and the proposed claims may be apparent in answering questions. • Claims are focusing on all the questions but loosely connected with the questions
4	• Strong Connection • Questions and claims fit strongly • Student develops claims for most of the generated questions

4	• Proposed claims are evident in answering questions even though claims are only for some of the generated questions
5	• Very Strong Connection • Questions and claims fit very strongly together • Student develops claims for all the generated questions and all the provided claims are obvious in answering questions

Scoring Matrix for Evidence

point	criteria
1	• Very weak evidence • Inaccurate, invalid, and unreliable evidence • Evidence is very sparse • Their observation is itself evidence (e.g., see my observation, calculation, or data section) • Evidence seems to come from no where in particular
2	• Weak evidence • May not be accurate, valid, and reliable • Evidence is simply a description of data • Evidence is from textbook
3	• Moderate evidence • May be valid evidence • May be reliable evidence • Evidence from textbook with a limited interpretation or explanation
4	• Powerful evidence • Accurate, valid evidence • Reliable evidence • Evidence from the interpretation of their collected observation/data
5	• Very powerful evidence • Very accurate, valid, rich evidence • Very credible and reliable evidence • Evidence from interpretation of their collected observation/data

Scoring Matrix for Claims-Evidence Relationship

point	criteria
1	• Very weak connection between claims and evidence • Evidence is not focusing on the claims at all • Claims without evidence or evidence without claims
2	• Weak connection • Evidence supports claims loosely or inadequately

2	• Student provides evidence for a few claims • Proposed evidence may not be apparent in supporting claims • Evidence is focusing on a few claims loosely
3	• Moderate connection • Evidence supports claims reasonably • Student provides evidence for some of the generated claims and proposed evidence may be apparent in supporting claims. • Evidence is focusing on all the claims but loosely connected with claims.
4	• Strong connection • Evidence supports claims strongly • Student provides evidence for most of the generated claims • Evidence is strongly supporting claims even though it is about some claims • Evidence is focusing on all the claims and strongly connected with claims
5	• Very strong connection • Evidence very strongly, effectively, and thoroughly supports the proposed claims • Student provides evidence for all the generated claims • Evidence is very strongly supporting all the claims

Scoring Matrix for Multiple Modal Representations

point	criteria
1	• Mono–modal representations or no representation • Only text
2	• Bi–modal representations • Text and graph • Text and math equations • Text and chemical equations • Text and diagram
3	• Tri–modal representations • Three kinds of representations
4	• Multiple–modal representations • Four kinds of representations
5	• Multiple–modal representations • Five kinds of representations • Examples: text, math representations, chemical representations, graph, tables, and diagrams

Scoring Matrix for Reflection

point	criteria
1	• Very weak explanation for why ideas have changed or have not changed • Student is not able to link to their own investigation to their existing knowledge • Student does not spot errors • Student does not have new questions
2	• Weak explanation for why ideas have changed or have not changed • Student may not be able to link their own investigation to their existing knowledge • Student may not spot errors • Student may not have new questions
3	• Moderate understanding of why ideas have changed or have not changed • Student may understand how their investigations tie into concepts about what they have learned in class • Student may make connections to concepts • Student may spot errors and may not explain them • Student may have new questions
4	• Strong understanding of why ideas have changed or have not changed • Student understands how their investigations tie into concepts about what they have learned in class • Student make some connections to concept and real life • Student spot errors and has some explanation for them • Student has new questions
5	• Thorough explanation for the change or lack of change in their idea • Student strongly understands how their investigations tie into concepts about what they have learned in class • Student refers some real life application to make a connection with their laboratory work • Student has suggestions for correcting their errors • Student recognizes what new things they have to think about • Student has new testable questions that are related to the investigation

Appendix B. Scoring Matrix for the Holistic Argument

point	criteria
2	• Very weak argument • Untestable questions, invalid claims, and unreliable evidence • Very weak connections between questions, claims, and evidence • Does not flow smoothly from one area to another
4	• Weak argument • May be untestable questions, invalid claims, and unreliable evidence • May not have reflection • Weak connection between questions, claims, and evidence • May not flow smoothly from one area to another
6	• Moderate argument • May be significant questions, adequate claims, appropriate evidence, and reflection • Moderate connections between questions, claims, and evidence • May flow smoothly from one area to another
8	• Powerful/enriched argument • Significant questions, valid claims, strong evidence, and meaningful reflection • Strong connection between questions, claim, and evidence • Flows smoothly from one area to another
10	• Very powerful/Enriched argument • Essential questions, very sound claims, very strong evidence, and very meaningful reflection • Very strong connection between questions, claims, and evidence • Flows very smoothly from one area to another