Understanding Student Reasoning in Writing:

Developing analytic frameworks for Biology and Engineering

An Honors Thesis for the Center for Interdisciplinary Studies

Liren Fu

Tufts University, 2018

Acknowledgements

To my thesis committee, thank you for your unending patience, understanding, and support for this project. To Dr. Julia Gouvea, for believing in me, for teaching me how to do research, and for providing the many hours of consultation, feedback, and encouragement necessary to turn this project into a reality. To Dr. Kristen Wendell, for the constant support, for always being a source of advice, and for always being willing to support my next research venture. To Dr. Modhumita Roy, for having mentored me throughout my Tufts career, for pushing me to grow, and for every word of wisdom. In many ways you are the three people most responsible for how I have grown as a student, researcher, and person in my time at Tufts, and I cannot thank you all enough for your influence and mentorship.

To the Tufts Summer Scholars Program, for providing invaluable funding for this project. To Dr. Aditi Wagh, Matt Simon, Robert Hayes, Brenna Gormally, and the Biology Learning Lab in general, thank you for all of the support you have given to this thesis, and for the support you have given me. To my housemates Charlotte Warne, Justine Aquino, Olivia Nkwonta, for every late-night word of encouragement as I stayed up analyzing data. To Ioana Tobos, for keeping me sane and always being here to hear me out. To Robert Middlemist, for constantly checking in, being here, and always reminding me of why I started this project. This thesis could not have happened without all of you, and I am eternally grateful for your presence in my life.

Introduction

Biology laboratory courses can engage students in deep scientific thinking, enhancing their reasoning and understanding of biological concepts through laboratory experiments (Sundberg & Moncada, 1994; Gasper & Gardner, 2013). This can be particularly evident in laboratory reports, with student writing having been shown to significantly improve critical thinking skills in Biology (Quitadamo & Kurtz, 2007), and being a place to practice scientific argumentation (Kuhn, 2010). However, there have been two obstacles in having laboratory reports engage students in scientific argumentation and reasoning.

Firstly, typical laboratory reports are often interpreted by students as exercises in reporting findings rather than making an argument (Peker & Wallace, 2009), or that the primary assessment criteria was "answering the question" instead of argumentation and reasoning (Zeegers & Giles, 1996). Secondly, assessing how students are reasoning in laboratory reports is a complex and interdisciplinary endeavor, with a multitude of assessment schemes having been proposed in the literature (Sampson & Clark, 2008). However, these assessment schemes can conflate language proficiency with student reasoning, through over-specifying the linguistic format and structure of reasoning (e.g. Lawson et al., 2000).

We conducted our study to investigate whether these two obstacles could be overcome, through explicitly reframing Biology laboratory report writing as about reasoning and argumentation as well as attempting to develop an analytic framework that minimized the effects of language proficiency in characterizing student reasoning. In addition, we were also interested in whether the analytic framework we developed could be applied to other disciplines, and thus attempted to apply it to Engineering design reports as well. From the results of a preliminary analysis, we suggest that there is promising evidence that simply reframing Biology laboratory reports as about demonstrating reasoning can lead to significant increases in demonstrated student reasoning. Moreover, we propose that we have moved closer towards developing an analytic framework which minimizes conflating reasoning and language proficiency, as well as towards a framework that can be applied across multiple disciplines. This thesis will first present the theoretical background of this project, then turn towards the work done with regard to Biology laboratory reports. It will subsequently consider the applicability of the analytic framework to Engineering design reports, and then discuss the findings as a whole.

Theoretical Background

We began this study from the perspective of a "resource model" of knowledge where students have important cognitive resources that they bring into play when reasoning (Hammer, 2000). Knowing that how students frame an activity influences how they engage in argumentation (Berland & Hammer, 2012), we suspected that there were richer reasoning structures that were simply not being captured or elicited in typical lab reports (e.g. Peker & Wallace, 2009), as opposed to prior research that has claimed students have trouble with argumentation (e.g. Kuhn, 2010). Thus, we wanted to investigate how explicitly reframing the purpose of lab reports could possibly elicit richer reasoning.

We also drew on Chomsky's distinction between linguistic performance and linguistic competence (1965) in framing how we understood analyzing student text. We propose that in trying to understand student reasoning from their writing, we are trying to understand a cognitive process that has been filtered through a dual layer of linguistic competence and performance in writing. As such, too strict a definition of reasoning, for example assessing the presence of hypothetico-deductive thinking only through the completion of an "if/and/then/but/therefore" linguistic format (Lawson et al., 2000), misses capturing all the ways that students are reasoning for evaluating linguistic performances that are not equivalent to reasoning in the first place. We thus sought to develop an analytic framework that could account for the variation and complexity in the lab reports we received, rather than measuring whether they were matching linguistic or epistemic structures that we had selected *a priori* as important (e.g. Sampson & Clark, 2008; Furtak et al., 2010).

We argue as well that since we are evaluating student reasoning many layers removed from the act itself, we can only truly assess what students have decided to demonstrate of their reasoning. In contrast to prior research, which has taken poor performance in research studies to indicate a lack of reasoning abilities (e.g. Sadler, 2004; McNeill et al., 2006), we propose that we can only make claims about the demonstration of students' reasoning abilities from what they perform, rather than the absence of these abilities. As such, this thesis is fundamentally trying to consider and understand changes in students' demonstrated reasoning, rather than claiming to be examining their cognitive abilities through a one-time response to a writing prompt.

Biology Laboratory Reports

Study Context

We conducted our study in a second semester introductory biology course with a three-hour weekly laboratory section, in a private research university in the North-eastern United States. For this paper, we are analyzing changes in the discussion section of the first lab report of the semester over two years, as that was the section of the lab report focused on student reasoning and argumentation. Students conducted the same experiment for the first lab report in both years: two strains of *E. coli* that differed in mutation rate were grown in a benign nutrient agar environment and a novel antibiotic environment, with rifampicin as the antibiotic. However, the structure of the course and the writing prompts were different.

In the original, "traditional", laboratory course in the first year, students read pre-lab materials and took a quiz at the start of the course. Students were guided to give a "Onesentence summary of what you found" and "a statement of how your results conform to your hypothesis" in the discussion section of their lab report. In the second year, we explicitly reframed the laboratory course, with students beginning with a group discussion on the benefits and costs of high and low mutation rates, as well as having access to a computer simulation to explore the effects of mutation rates beyond the experiment they set up. In contrast to the first year, students were asked in the guidelines to reason about "What does the experiment tell you about some of the questions raised in this lab, and what does the experiment *not* tell you?" in the discussion section. Additionally, they were informed in the guidelines that "You will be graded for the logical flow and for evidence of your own thinking", and "Don't be afraid to say things that are "incorrect", but be sure to fully explain your thinking".

Methods and Analytic Framework

We limited our analysis to a subsample of lab reports from each year (24 from the first year, 19 from the second), limited to those from sections taught by the same graduate TA over the two years. We initially developed a preliminary framework based on an analysis of "epistemic levels" proposed by Kelly and Takao (2002), where they used an analytic model that considered both the disciplinary-specific knowledge and theoretical generality of the propositions found in student writing in an introductory oceanography course to assign epistemic levels, with "I" representing the most grounded propositions explicitly referencing data charts and "VI" representing the most general propositions referencing geological processes. However, while our analytic framework still draws heavily from their work, our attempts to replicate their epistemic levels surfaced methodological problems when applied to Biology. While Kelly and Takao's model thought of geological generality as a proposition that could be applied to larger and larger geological areas, we found no adequate corollary in

Biology, since propositions about cellular dynamics can be as general or as specific as claims about organismal fitness.

Since students could make claims about fitness and survival at the level of individual colonies on an agar plate, to a particular strain of *E. coli*, to a particular genotype expressed by multiple strains of *E. coli*, to all organisms expressing a high or low mutation rate, we decided to shift away from their model towards an emergent coding scheme instead. We looked for factors that emerged as important qualitative differences within lab reports, ultimately developing three primary coding schemes that we will be presenting in this thesis. Recognizing the tension between wanting to reduce the effects of language proficiency on our analysis of fundamentally linguistic texts, we based these coding schemes on flexible linguistic divisions and markers that we took as indicative of significant factors of reasoning. These linguistic divisions and markers could be single words, or phrases, or syntactic in nature, or contextual cues found in other parts of the text, or any combination of the above. In allowing for multiple indicators for any factor of reasoning as well as contextual cues, we aimed to develop an analytic framework that could interpret the variety present in students' linguistic competences and performances with as much sensitivity as possible.

The first coding scheme we developed was based on noting what linguistic divisions were emerging as meaningful in our analysis of the data, resulting in what we are calling argument units (Table 1). While the scheme was initially restricted to coding T-units (Hunt, 1965), which are "main clause[s] with all subordinate clauses attached to [them]", this would result in ambiguous situations where two or more codes could apply. For example, the T-unit "both strains grown without rifampicin formed a lawn because the conditions were very favorable" would simultaneously qualify as both a piece of evidence ("formed a lawn") and an explanation ("because"), and thus the scheme was modified to accept clauses and other linguistic units as well.

Table 1. Argument Units for Analysis of Biology Laboratory Reports: Definitions and Examples

Category	Definition	Examples
Question	Any T-unit/clause marked as a question.	"…then why is DNA repair so pervasive?"
Goal	Any T-unit/clause marked as a desired outcome through words like "want" or phrases like "in order to".	"…in order to keep the total amount of bacteria in an observable range…"
Claim	Default code, applied to T-units that no other code applies to, as well as fragments and clauses left between other codes and generated linkages.	"Some of these mutations may have been deleterious…"
Explanation	Any T-unit/clause beginning with a "because" synonym, or marked linguistically by the text as an explanation.	"Because the E398 strain had suboptimal DNA repair mechanisms…"
Evidence	Noun, noun phrase, clause, or T-unit relating to observable experimental evidence or statistical trends.	"…the numbers of colonies grown are much lower."

Elements of this coding scheme are analogous to assessment schemes that have previously been proposed in the literature; in particular, what we are calling Claim-Evidence-Explanation is analogous to the Claim-Evidence-Reasoning distinction that has been proposed by others (e.g. McNeill et al., 2006; Furtak et al., 2010), which is itself a modification of Toulmin's (1958) framework. While we acknowledge the many similarities between our Claim-Evidence-Explanation coding and the Claim-Evidence-Reasoning distinction, particularly as used by Furtak et al. (2010), we do want to draw attention to two differences. The first is that we are reframing what has previously been called Reasoning as an Explanation instead; while this may seem a pedantic distinction, we argue that it is important enough to reframe, as if reasoning is what we as a field are interested in capturing, then are we simply asking for students to increase their usage of the word "because" in order to increase the amount of "Reasoning" they are doing? While we do think it is important to capture justificatory statements made by students as they posit a particular proposition as a causal mechanism for another, we do not think these statements are necessarily analogous or equivalent to reasoning, which can consist of much more complicated argument chains.

Secondly, we found in our analysis two additional codes of Goals and Questions. Goals marked a particular proposition as something to be fulfilled or achieved, while Questions marked something as theoretically unresolved; these complimented the Claim-Evidence-Explanation framework, with Claims tending to be disciplinary propositions, Evidence being scientific data, and Explanations being the justificatory disciplinary backings proposed for other argument units.

Table 2. Argument Modes for Analysis of Argument Units: Definitions and Examples

Category	Definition	Examples
Hypothetical	Any argument unit marked as a hypothetical or occurring in the future.	"For future experiments, E. coli could be put in many different environments to test how its mutations help it to survive…"
Unmarked	Default code, applied to any argument unit that no other code applies to	"…because it was able to mutate faster to gain antibiotic resistance."
Referenced	Any argument unit with information from the pre-lab readings, or which could not have come from the experimental space, or cited, and not marked as an inference from another argument unit	"Most eukaryotic cells have functional systems of checkpoints and proofreading to avoid flawed replicated DNA."
Simulation	Any argument unit marked as about the simulation space provided in Year 2 of the study	"In the simulation, adding poison to the environment resulted in more yellow bacteria…"

Based on our analysis of the argument units we found, we developed a second coding scheme to be used in conjunction with it (Table 2). This coding scheme noted whether argument units were marked as being hypotheticals (most often hypotheses or future experimental ideas), references (with information from sources other than the physical experimental setup), about the simulation provided in the second year, or simply unmarked. This helped provide additional context to the argument units, allowing us to look at what mode of information they were working with, helping to make qualitative differences between engaging in different theoretical spaces. These two coding schemes together allowed us to do quantitative analyses on the differences in proportions of argument units and modes between the two years, identifying coarse-grained signals about possible differences in elicited student reasoning from the different question prompts.

Table 3. Generating Argument Unit Linkages for Analysis of Biology Laboratory Reports: Definitions and Examples

Category	Definition	Examples
Supporting	Default linkage generated between two units	"…further supports…"
Opposing	Generated when two units are marked as indicative of opposing things, or when contextually opposite	"…however…"
Repetition	Generated when two units are the same, though occurring at different points in the report	"higher numbers of E938 mutant colonies in the Rifampicin environment…the high numbers of mutant E938 colonies in the Rifampicin environment…"
Functional Grouping	Generated when a phrase referenced multiple argument units as a group	"All the reasoning behind…"

However, we were also interested in Kelly and Takao's (2002) proposal that the different propositions in student writing could be linked to each other, thus generating a model of the students' argument structure. As such, we developed a third coding scheme (Table 3) for modelling the linkages between the different argument units we coded, creating chains of argument units that represented the complexity of student reasoning in the reports. We defined linkages between different argument units as stemming from the presence of a conjunction or conjunction phrase, the repetition of a noun or noun phrase or idea, or the presence of syntactic markers like commas or semi-colons. We did not model linkages between each argument unit and every other argument unit that shared a noun or noun phrase or idea with it; rather, we generated linkages between each new argument unit that was coded and previous chains that contained those shared units, thus preferentially generating linear argument chains.

This allowed us to model the argument structures that students were generating in their lab reports; the different types of linkages also allowed us to consider how different units were being used in context, tracking which parts of the laboratory reports were supporting each other and which ones represented a different view. As each argument mode code was tied to an argument unit code, and linkages between the different argument units were also coded, we could also use the same linkage structure for representing changes in argument modes, or for any other coding scheme we wished to develop that would follow the boundaries of the argument units. Furthermore, as each argument unit occurs sequentially in a laboratory report, we could assign numbers to each argument unit, thus allowing us to order them against each coding scheme in order to generate graphical representations of students' argument structures. As the argument mode scheme is dependent on the argument unit scheme, the two could also be layered onto each other to observe interaction effects, thus allowing us to attend to the structure of each report in greater qualitative detail (Figure 1), as well as to changes in how students were linking argument units and modes from year to year.

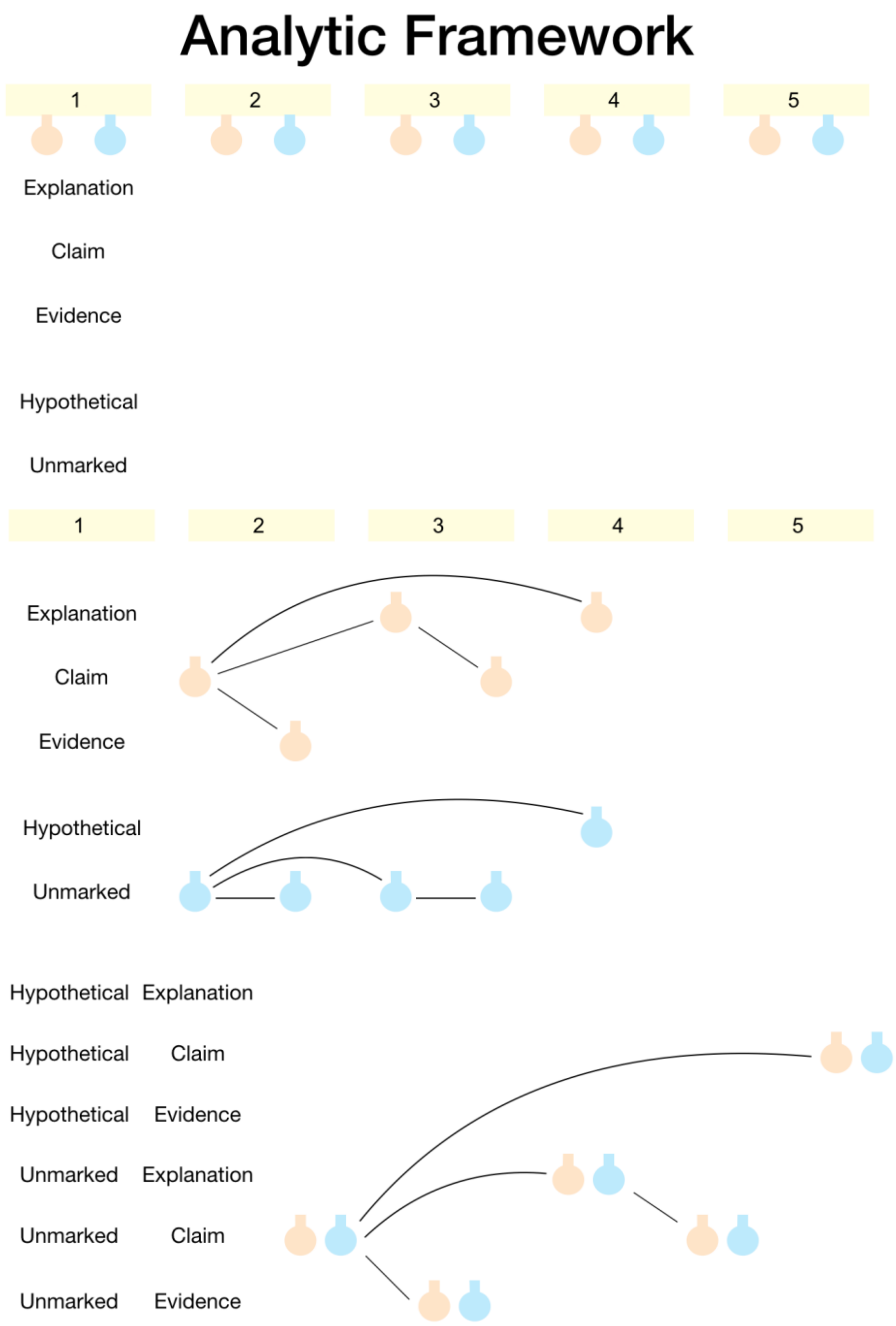

Figure 1. Graphical representation of analytic framework modelling. Tags in orange and blue represent argument unit and mode codes, respectively, being graphed in the order they occur on the x-axis against their respective coding schemes, and respective coding schemes layered onto each other, on the y-axis.

As we were also interested in the content of what students were reasoning about, we developed a fourth coding scheme to capture the actual substance of what students were reasoning about (Table 4). As reasoning content was defined at a much smaller grain size than either the argument unit or argument coding scheme, with most reasoning content being fragments of a clause, this allowed us to have a finer-grained analysis of what students were reasoning about within their larger argument structures. As reasoning content was coded within the boundaries of each argument unit, similarly to argument modes, we could theoretically layer it onto the larger argument structure; however, as each argument unit has a variable amount of reasoning content, we faced methodological problems in terms of representing that amount of data on the structure, and so did not pursue that avenue of analysis.

Table 4. Reasoning Content for Analysis of Biology Laboratory Reports: Definitions and Examples

Category	Definition	Examples
Genetic	Any idea involving consideration of the genotype, including genetic changes and possession of particular genes	"…started with lac- gene…"
Parametric	Any idea involving the rate or probability of some event happening	"…increased mutation rate…"
Organismal	Any idea involving organismal traits, such as strain or expression of a phenotype	"…the use of lactose…"
Population	Any idea involving population values such as fitness or lifespan	"…the long-term survival…"
Environmental	Any idea involving environmental factors such as temperature or resource availability	"…under the environment of antibiotics…"
Medical	Any idea involving the field of medicine	"…the drugs that humans use to combat them."
Experimental Methods	Any idea involving the experimental set-up or procedures	"…plating a smaller amount of cells…"
Experimental Data	Any idea involving experimental data	"…count of colonies…"

Out of the eight categories that were generated, we were most interested in the first six, as opposed to the last two categories, Experimental Methods and Experimental Data. This is as we wanted students to reason about what the experiment could tell them about biological concepts and to posit biological theories for the data they were generating in their experiments, rather than simply accounting for unexpected data as a result of experimental error in procedures or providing large amounts of data without reasoning about what the data meant in terms of biological theories. As such, the first six categories were additionally tagged as Biology Conceptual Content.

We were also interested in whether the Biology conceptual content students were providing in their laboratory reports matched what had been provided in the pre-laboratory readings, or were nuanced differently from the pre-laboratory readings, or were not to be found in the pre-laboratory readings at all. As such, we compared all argument units with Biology conceptual content against information provided in the respective year's pre-laboratory reading in order to establish whether this content had been given to students. When all of the Biology conceptual content in an argument unit matched what was given in pre-laboratory readings, they were additionally tagged as Given Content; however, there were cases where some of the conceptual content in an argument unit would match what was given and some would not, and we would tag the conceptual content which matched as Given while leaving the rest untagged (Figure 2). This allowed us to better capture what reasoning content was actually provided in pre-laboratory readings, rather than coding all of the content in a given argument unit as given or not given.

In this thesis, we'll be presenting two avenues of analysis we pursued in order to determine whether there were changes in the lab reports between the two years. The first avenue of analysis we pursued was whether there were changes in how students were using and integrating experimental data into their laboratory reports. In order to examine this, we looked at the proportion of Evidence units out of the total count of argument units in each laboratory report, as well as the proportion of Evidence-Evidence linkages out of the total count of linkages, to get a sense of whether students were simply providing large amounts of data that were not being reasoned about or integrating the pieces of experimental data they had generated with the theoretical claims they were making. The second avenue of analysis was whether there were changes in the amount of Biology conceptual content being provided, and changes in the amount of that conceptual content that was given in pre-laboratory readings. We analyzed this through looking at the proportion of reasoning content units coded as Biological Conceptual Content out of the total count of content units for each laboratory report, and the proportion of conceptual content that was coded as Given out of the total count of Biological conceptual content. We used SPSS Statistics Version 24 to run quantitative analyses and ran independent samples T-tests to check for statistical significances.

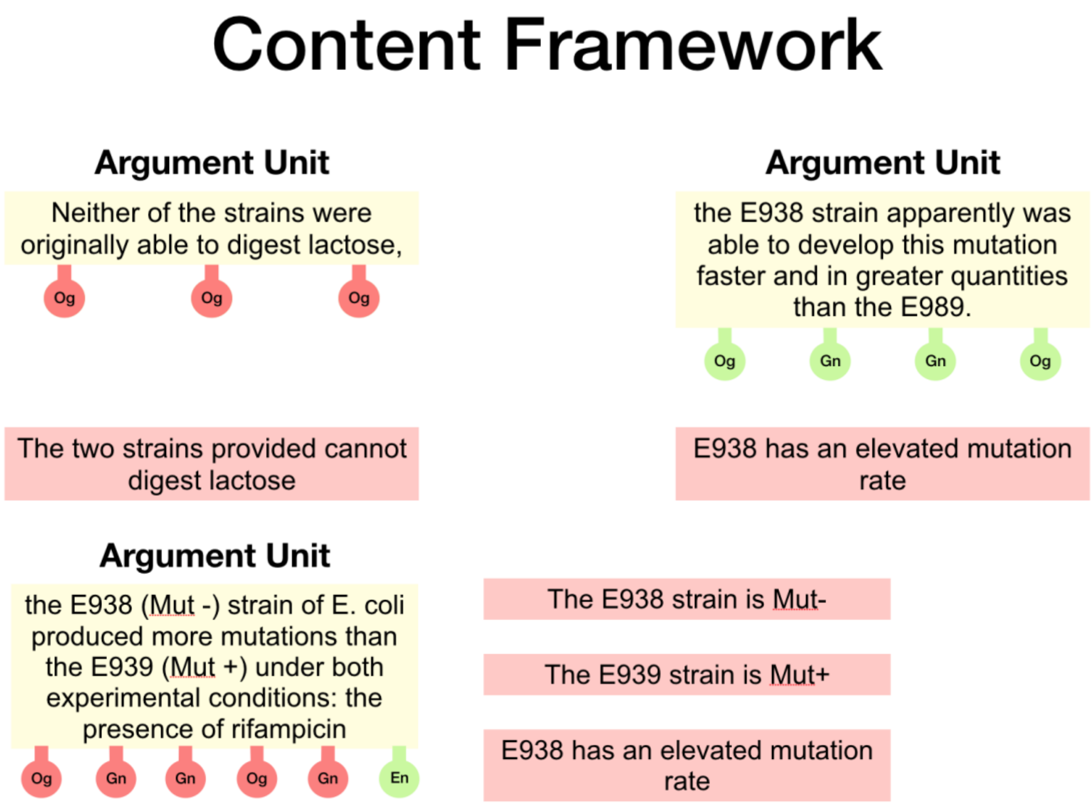

Figure 2. Graphical representation of analytic framework content coding. Tags in red and green represent given Biological conceptual content and non-given Biological conceptual

content codes, respectively. Texts in yellow are argument units from laboratory reports, while texts in red are the information provided in pre-laboratory readings that the argument units are being compared against to decide if there is a match.

Findings

<u>Changes in Evidence Usage</u>

Table 5. Mean Percentage of Evidence argument units in laboratory reports in the two years.

Statistic	Year 1	Year 2
Number of Reports Analyzed	24	19
Average Percentage (%)	24.6	24.7
Standard Deviation (%)	12.7	10.4

Based on a preliminary application of the argument unit coding scheme to the whole data set, there was no change in the average proportion of Evidence argument units between the two years (Table 5), indicating that students were using about the same proportion of evidence in Year 2 of the study as in Year 1. However, as there were significantly more average argument units (Table 6, $p < 0.01$) and total word count (Table 7, $p < 0.01$) per laboratory report, students were ultimately providing more pieces of evidence in Year 2 as opposed to Year 1.

Table 6. Mean Count of Argument Units in laboratory reports in the two years.

Statistic	Year 1	Year 2
Number of Reports Analyzed	24	19
Average Argument Units	24.5	40.2
Standard Deviation	8.7	16.6

Significance	$p < 0.01$

Table 7. Mean Total Word Count of laboratory reports in the two years.

Statistic	Year 1	Year 2
Number of Reports Analyzed	24	19
Average Word Count	317.8	487.6
Standard Deviation	123.6	193.2
Significance	$p < 0.01$	

When we analyzed the average proportion of Evidence-Evidence linkages in the laboratory reports, we found a non-significant decrease in the average proportion in Year 2 even after outlier values were removed (Table 8, $p < 0.1$). This suggested that though students were integrating the pieces of evidence they were providing in Year 2 more with other kinds of argument units, rather than simply providing long chains of evidence, there was not a significant change in the number of students actually doing this. Analyzing the scatter plot of the reports analyzed, while there were more laboratory reports in Year 1 than Year 2 which had >20% Evidence-Evidence linkages, there was a substantial number of laboratory reports with >20% Evidence-Evidence linkages in both years (Figure 3). Taking both trends together, though there are some changes in how students are using evidence in the two years, the changes are not particularly significant, suggesting that the difference in reframing did not really affect student usage of evidence.

Table 8. Mean Percentage of Evidence-Evidence linkages in laboratory reports in the two years.

Statistic	Year 1	Year 2

Number of Reports Analyzed (Outliers Removed)	24	17
Average Percentage (%)	13.3	8.0
Standard Deviation (%)	12.5	5.3
Significance	$p < 0.1$	

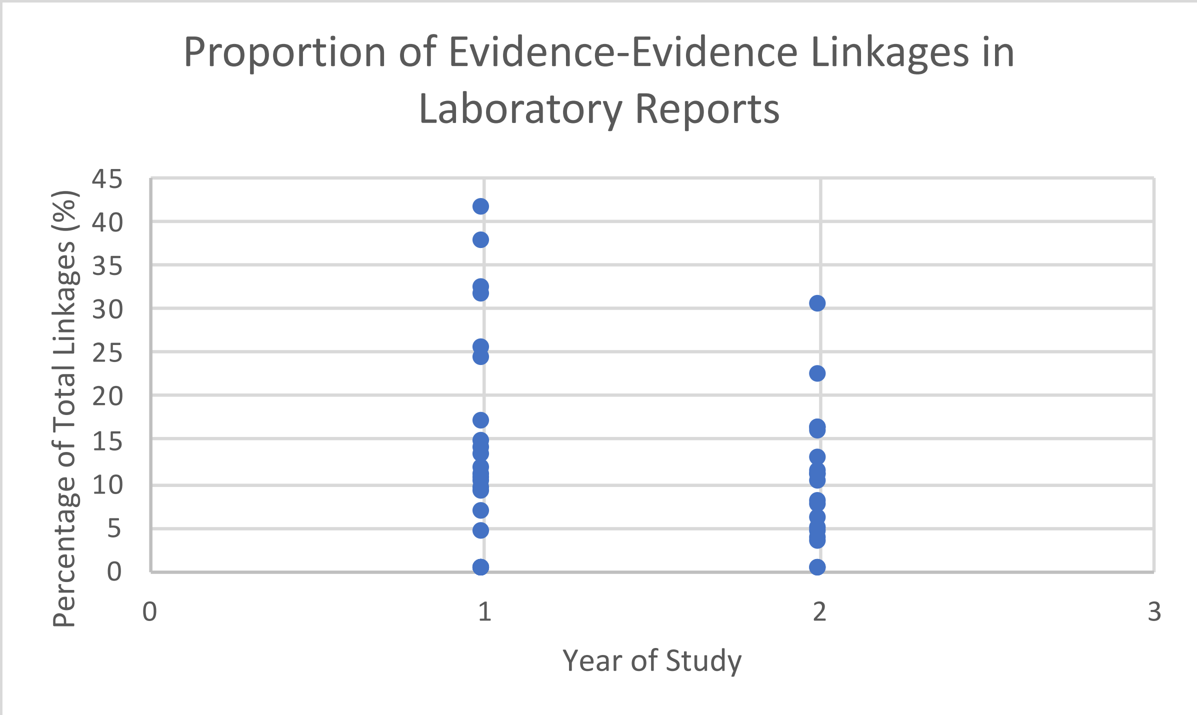

Figure 3. Scatter plot of proportions of Evidence-Evidence linkages in laboratory reports in the two years of the study. 24 reports were analyzed in Year 1 and 17 reports in Year 2.

<u>Changes in Reasoning Content</u>

Table 9. Mean Percentage of Biology Conceptual Content in laboratory reports in the two years.

Statistic	Year 1	Year 2
Number of Reports Analyzed	12	10

Average Percentage (%)	73.6	80.9
Standard Deviation (%)	6.6	9.1
Significance	$p < 0.05$	

Based on a preliminary application of the reasoning content coding scheme to a subsample of the whole data set, consisting of 12 reports from Year 1 and 10 reports from Year 2, the proportion of Biological conceptual content significantly went up in Year 2 (Table 9, $p < 0.05$), suggesting that the reframing of the discussion section prompts led to students positing more theoretical claims rather than simply providing content about experimental methodologies or data. When analyzing the proportion of given conceptual content, we found that it dropped very significantly in Year 2 (Table 10, $p < 0.01$), and when looking at the scatter plot (Figure 4), we note that all laboratory reports in Year 2 had less than 20% given conceptual content whereas there were multiple laboratory reports in Year 1 with more than 20% given conceptual content. This tentatively indicates that the reframing resulted in students providing more nuanced and novel propositions about Biology concepts, rather than simply repeating ideas that had already been provided to them from the pre-laboratory readings.

Table 10. Mean Percentage of Given Biology Conceptual Content in laboratory reports in the two years.

Statistic	Year 1	Year 2
Number of Reports Analyzed (Outliers Removed)	11	10
Average Percentage (%)	25.7	11.9
Standard Deviation (%)	13.6	5.5
Significance	$p < 0.01$	

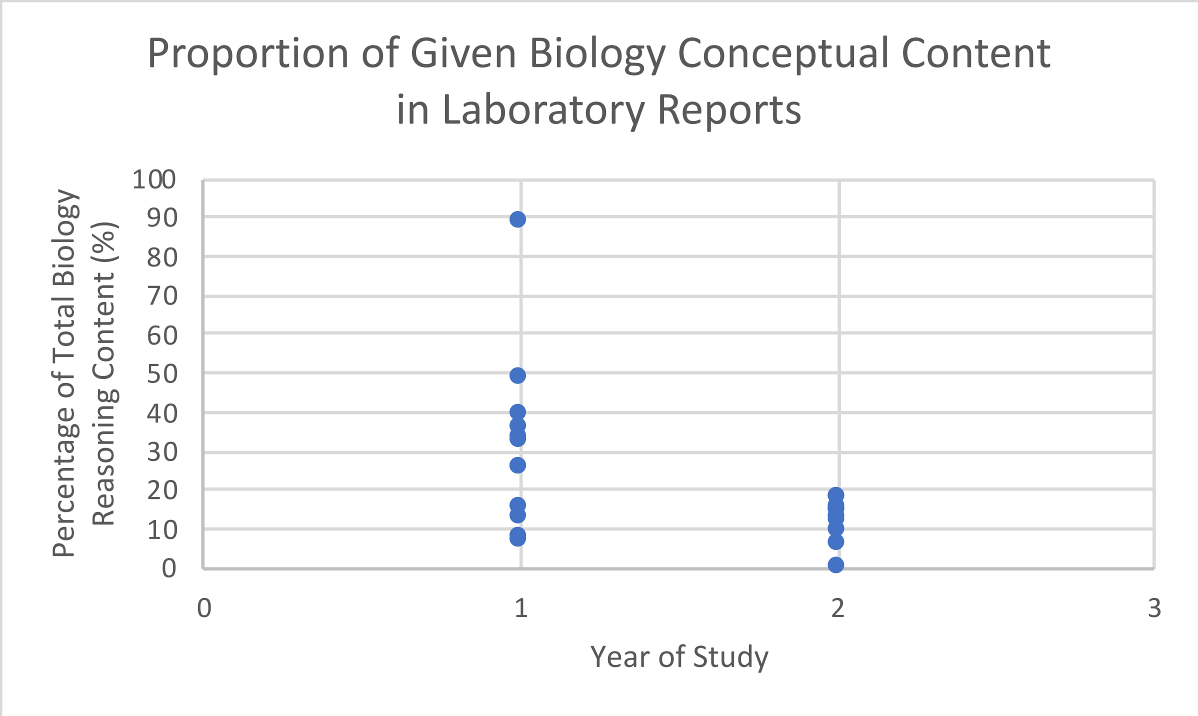

Figure 4. Scatter plot of proportions of Given Biology Conceptual Content in laboratory reports in the two years of the study. 12 reports were analyzed in Year 1 and 10 reports in Year 2.

While these are promising initial results, suggesting that the reframing of the laboratory course structure resulted in students both providing more conceptual claims and Biological theories, as well as not simply repeating what had been provided in pre-laboratory readings, we do note that this is still a preliminary analysis of a subsample, which needs to be applied to the whole data set and run through inter-rater reliability in order for us to make these claims with confidence.

Understanding Changes in Context

While the quantitative statistics provide a useful coarse-grained signal for understanding differences between the two years broadly, there is still the question of what these statistics, like the proportion of Evidence-Evidence linkages, mean in terms of actual writing in the laboratory reports. What we aim to do in this section of the thesis is to attempt to tie the different quantitative signals together to show what their numerical values actually translate to in the form of the actual text of the laboratory reports and our qualitative senses of them. We will be presenting three different laboratory reports in this section, providing both their quantitative statistics as generated through our analysis and their modelled argument structures, and connecting these analytical values to the actual text of the report, in order to demonstrate the meaning of these signals in the context of the laboratory reports and provide additional nuance to what these values could indicate.

Table 11. Quantitative Data from evidence usage and reasoning content analysis of all reports from both years, 3AS23, 3AS11, and 3BS6. One outlier value was removed from the average proportion of given conceptual content. 43 laboratory reports were analyzed to generate the average.

Statistic	Average (Year 1 and 2)	3AS23	3AS11	3BS6
Total Word Count	392.8	328	319	406
Total Argument Units	31.4	24	24	27
Evidence Argument Units (%)	24.7	41.6	33.3	25.9
Evidence-Evidence Linkages (%)	11.8	32	24	7.4
Biological Conceptual Content (%)	76.9	75	63.9	88.8
Given Conceptual Content (%)	19.1	25.9	33.3	12.7

To begin, we will consider 3AS23, a report from the first year, which has both a high proportion of Evidence argument units, a high proportion of Evidence-Evidence linkages, an approximately average proportion of Biological conceptual content, and an above average proportion of given conceptual content, compared to the average proportion of the respective values in both years (Table 11). Looking at its modelled argument structure, it has a string of eight Evidence units in the middle of the laboratory report, which is then connected to a chain of reasoning derived from the pre-laboratory reading (Figure 5).

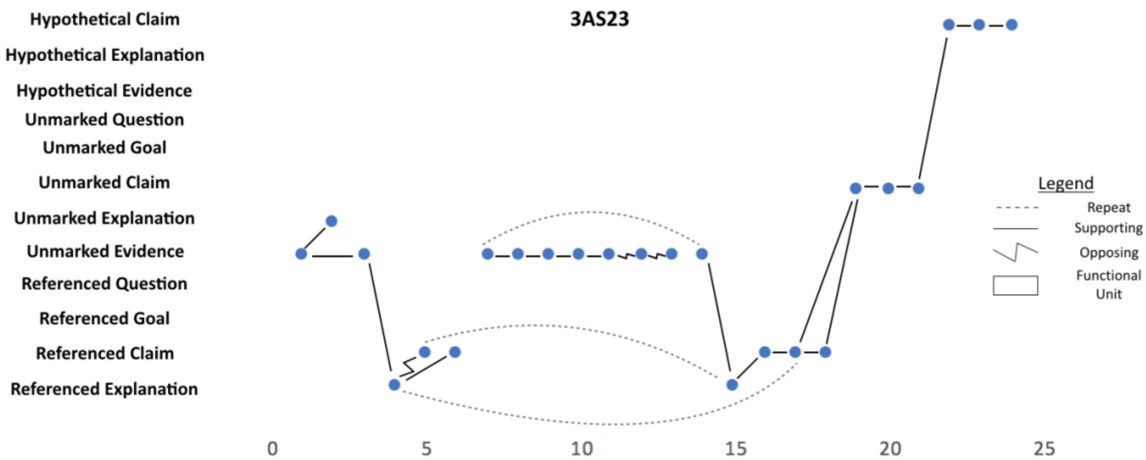

Figure 5. Modelled argument structure of laboratory report 3AS23, from the first year of the study.

This would suggest that rather than reasoning about each piece of evidence they were providing or using them intentionally to support novel and nuanced theoretical propositions, this student in the first year of the study was simply presenting their experimental data as the bulk of the report, possibly in alignment with Peker and Wallace's (2009) finding that students interpreted typical laboratory reports as an exercise in reporting findings. This is validated by looking at the text of the laboratory report (Figure 6): a long sequence of experimental data is provided, but only one piece of data is explicitly reasoned about by the student in the text, with information that matches what was already provided as a given in the pre-laboratory readings.

In the adaptation experiment, **strain E939 showed the highest percentage of trials lacking red papillae**, and **strain E938 had the highest percentage with red papillae**. **E939's percentage of runs with no papillae present was about triple that of E938's**. **The results of both strains in the 1-2 papillae category were not statistically significant**, since their **standard deviations were too large**. **E939, however, had no cases of more than 2 red papillae**, whereas **E938 had almost half (Figure 1)**. **E939's low percentage of trials with papillae present** resulted from **the strain's functional DNA repair mechanisms**. These mechanisms meant that it produced mutations at a much slower rate than E938, which had suboptimal DNA repair mechanisms (Mut-).

Figure 6. Passage from laboratory report 3AS23, from the first year of the study. Text in pink was coded as Evidence, while text in blue was coded as Explanation, and text in green was coded as Claim.

In contrast, 3AS11, another report from the first year, has a comparatively lower proportion of Evidence argument units and Evidence-Evidence linkages (Table 11). This suggests that rather than simply providing large amounts of experimental data that are not reasoned about in the laboratory report, the student is making more claims or providing more explanations as compared to 3AS23, and the pieces of evidence that they are providing are more integrated with the claims and explanations that they are making and providing, thus acting as experimental support for chains of reasoning. However, it also has a lower proportion of Biological conceptual content and a higher percentage of given conceptual content (Table 11). This indicates that even though there are more claims and explanations, these claims and explanations are largely about the experimental methodology or match what has already been provided in the pre-laboratory readings, instead of providing novel or nuanced theoretical propositions.

Considering its modelled argument structure, we can also note a few interesting things. There is one central Hypothetical Claim that seems to act as a coordinating argument unit for the structure as a whole, since the rest of the argument branches out from it; there are also multiple argument linkages that connect argument units that are far apart in the actual report, with the supporting linkage between argument unit 1 and 24 being the most drastic example of this (Figure 7). This would suggest that rather than having one singular argument thread in the laboratory report, where each successive argument unit builds on the ideas expressed in the previous one, the laboratory report generates new argument branches based off older propositions, and the whole argument is ultimately based around supporting one particular Hypothetical Claim.

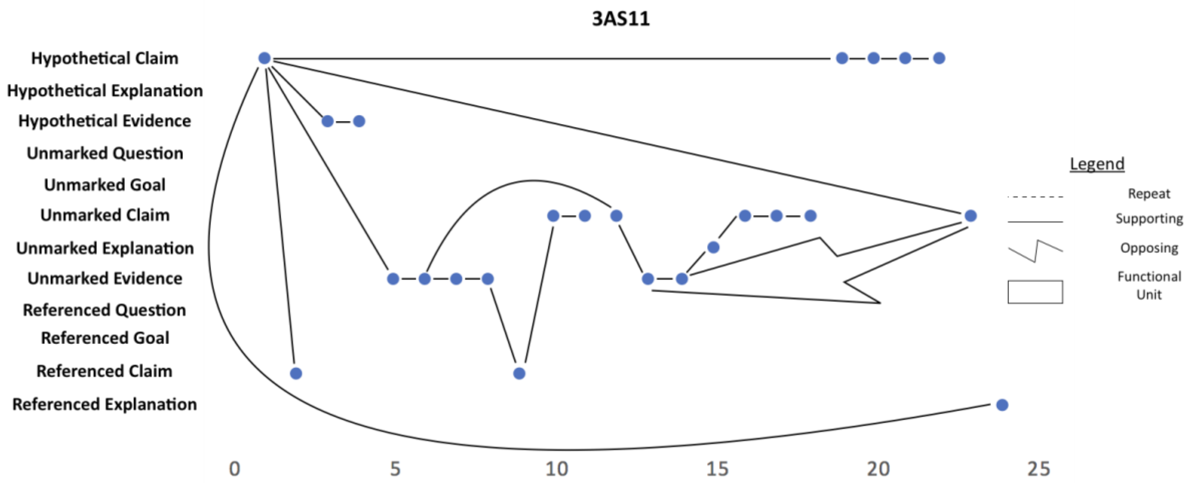

Figure 7. Modelled argument structure of laboratory report 3AS11, from the first year of the study.

This is validated by looking at the text of the laboratory report (Figure 8). The laboratory report begins with its hypothesis, which though posited as hypothetical completely matches the information provided in the pre-laboratory readings as a guaranteed outcome. This hypothesis is also linked to three separate argument chains, which are that E938 is Mut-, that it would be visualized by particular hypothesized evidence, and that it was supported by the actual experimental evidence, thus acting as a coordinating claim for this whole beginning passage. There is a developed chain of reasoning about the data provided, where the student is making claims about survival and fitness, which is then abandoned for elaborating on a particular piece of evidence provided previously. Both the text of the laboratory report and its modelled argument structure thus suggest that the student was trying to provide "how [their] results conform to [their] hypothesis", providing some initial evidence of the prompt affecting their framing and demonstrated structure of reasoning.

> Our hypothesis was that E938, [which is Mut-], would show increased mutation over strain E939. This would be visualized by **the E938's growth of more colonies on the agar with rifampicin** and **more red papillae on the MacConkey agar than the E939 strain with an intact DNA repair mechanism**. This hypothesis was supported by both **our individual** and **class data**, which showed **increased growth of the E938 over the E939 on rifampicin treated plates**. **E938 also developed more red papillae per colony on the MacConkey agar than did the E939**.
>
> The increased mutation rate and survival of E938 can be attributed to the faulty DNA repair of the Mut- strain. Increased mutation rate allows more bacteria of this strain to gain the traits necessary for survival on the rifampicin plates. Likewise, E938 also adapts better for the use of lactose as an energy source on the MacConkey agar. Some uncertainty was present in the class data for the E938 strain.

Figure 8. Beginning passage from laboratory report 3AS11, from the first year of the study. Text in pink was coded as Evidence and text in green was coded as Claim. Text in square brackets was coded as a separate argument unit from the text surrounding it.

Finally, to turn to 3BS6, a report from the second year, we see how the different quantitative signals in tandem can indicate a laboratory report with high demonstrated reasoning. 3BS6 has a roughly average proportion of Evidence argument units, a below average proportion of Evidence-Evidence linkages, an above average proportion of Biological conceptual content, and a below average proportion of given conceptual content (Table 11). This suggests that in contrast to the previous two laboratory reports, 3BS6 both integrates experimental data with theoretical propositions and provides theoretical propositions which are novel and nuanced and about Biology concepts rather than simply linking everything to experimental methodology. Considering its modelled argument structure, we can also note that all of the argument units can be traced in a single unbroken line, suggesting that there is only one large, cohesive argument that the student is consistently developing throughout the whole report (Figure 9).

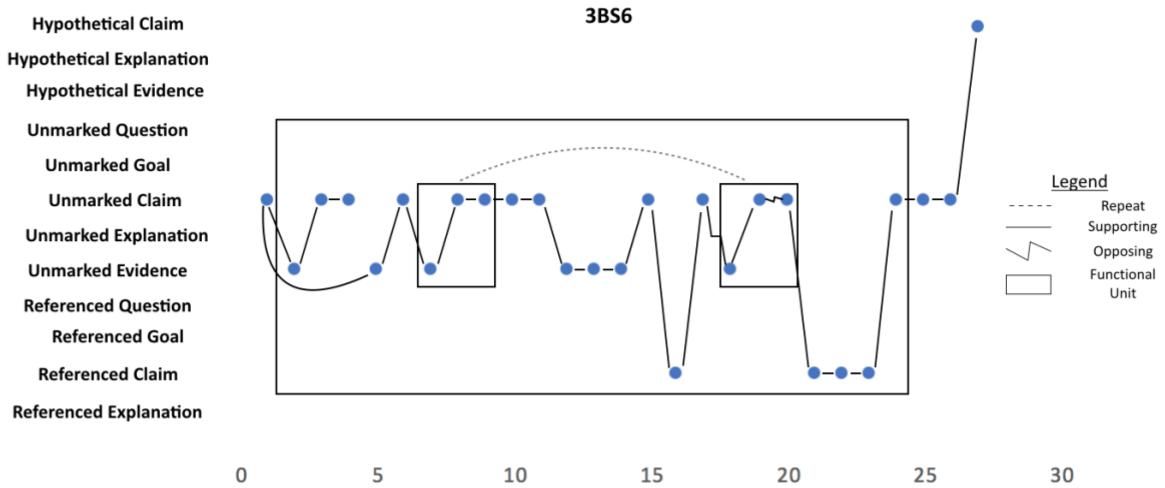

Figure 9. Modelled argument structure of laboratory report 3BS6, from the second year of the study.

This is again validated by looking at the text of the laboratory report (Figure 10). The student moves between pieces of evidence and strings of claims, integrating experimental evidence as support for novel and nuanced theoretical propositions which were not provided as givens. There are also a number of different ideas provided, from the comparative rate of specific beneficial mutations to survival and fitness in particular environments. Thus, in comparison to the previous two reports from the first year, this laboratory report demonstrates more interesting and sophisticated reasoning in terms of providing a variety of ideas that were not provided as guarantees in the experimental space and building on previous theoretical propositions with each new claim and piece of evidence. This provides some evidence that the reframing of the laboratory sections could influence how students choose to demonstrate their reasoning, shifting away from simply trying to defend their hypothesis to reasoning about the theoretical propositions that they could generate from the experimental data they had gathered.

The fact that **more mutant colonies grew on rifampicin than wildtype colonies** suggests that the higher mutation rate allowed more of the E. coli to develop a resistance to the antibiotic faster than the wildtype E. coli. **Some of the wildtype E. coli cells did survive the rifampicin**, suggesting that they had the same mutation, just in fewer numbers. In this particular environment, having a high mutation rate gave the E938 E. coli a distinct advantage over their E989 counterparts.

In the MacConkey agar plates, the E938 mutants also had a large advantage over the wildtype strain simply **in numbers of colonies regardless of color**. The **mutant colonies also had very significantly larger numbers of red colored spots** and **colonies than the wildtype colonies**. This indicates that the E938 strain colonies digested much more lactose than the E989 strains did. Neither of the strains were originally able to digest lactose, so the E938 strain apparently was able to develop this mutation faster and in greater quantities than the E989.

Figure 10. Passage from laboratory report 3BS6, from the second year of the study. Text in pink was coded as Evidence and text in green was coded as Claim.

These example laboratory reports thus provide both a defense of the usage of the analytic framework as well as a cautionary note. While the framework can generate useful, coarse-grained signals to indicate changes across different groups of laboratory reports, there is no one signal that is the definitive sign of "good reasoning" in laboratory reports. Instead, we argue that though each signal provides us some information, well-reasoned laboratory reports ultimately showcase a variety of signals simultaneously, which must be taken into account in context in order to figure out what is truly happening in terms of student reasoning in Biology laboratory reports.

Engineering Design Reports

Study Context

In order to investigate whether this analytic framework could be applied to other disciplines, we began an exploratory attempt at analyzing Engineering design reports from a senior capstone course in Mechanical Engineering in the same private research university in the North-eastern United States. However, while the Biology laboratory reports were written by individual students, these Engineering design reports were written by students in groups of 3 to 4. We identified three sections within the Engineering design reports to analyze, similar to how we focused on the discussion section of the Biology laboratory reports, as well as their respective writing prompts within the larger set of instructions for the Engineering design report: the Executive Summary, the Final Design Description, and the Revisions to the Design Solution. These three sections were chosen both for predicted similarities and dissimilarities to the analyzed discussion sections of the Biology laboratory reports. In the Executive Summary, the prompt was for "the important results conveyed in the report be summarized, while the data and the arguments which support them remain in the main body of the report". In contrast, the prompt for the Final Design Description asked students to "remember to describe, explain, and justify", and the prompt for the Revisions to the Design solution simply asked for "recommendations for the next version".

We hypothesized that if any section was to look most similar to the structures generated in Biology laboratory reports, it would be the Final Design Description, as the writing prompts were similarly focused on trying to "describe" and "explain" what had happened in students' experiences. We also predicted that the Revisions to the Design Solution would be similar to the hypothesized future experiment argument fragments in the Biology laboratory reports, which tend to be a chain of claims, as the prompts for both essentially ask for the same thing. Lastly, we suspected that the Executive Summary, in calling for "important results conveyed", might reveal important disciplinary differences between Engineering and Biology.

Methods and Analytic Framework

As this was a preliminary exploratory investigation into the multidisciplinary applicability of the analytic framework, we only analyzed sections from 1 Engineering design report, with the report having been chosen at random from the total sample of 7 Engineering design reports. In attempting to evaluate whether the analytic framework, as developed for Biology, could be applied to other disciplines, we chose to make very few modifications in applying it to the Engineering design reports in order to test whether it would still work. However, two changes did need to be made. The first was that the reasoning content analysis was not attempted as the coder was not an engineer by training, and thus did not presume to know enough to be able to generate appropriate and adequate categories for the disciplinary content displayed in an Engineering design report.

The second was that a change did need to be made in the argument unit coding scheme, due to methodological problems in deciding what counted as Engineering evidence that were surfaced by the same lack of domain knowledge expertise. While Biology draws a fairly sharp distinction between observable events as pieces of evidence and disciplinary theorizing as the subject of claims, the distinction was not as easily drawn in Engineering. This is due to the fact that in Engineering, the primary unit of output is physical, whereas in Biology, especially in introductory Biology laboratory courses, the primary unit of output is not a physical product but a theoretical proposition. As such, the coder made the decision to classify Engineering evidence as instances of successful trials of the design, or clauses or statements that could be interpreted as such. This necessarily brought a context-sensitivity to the decision-making process that did not have a corollary in Biology, where for example all instances of colony counts could be classified as pieces of evidence without reference to any other argument units.

While the analytic framework was still able to be applied, this did generate theoretical issues that require resolution in consultation with an Engineering disciplinary specialist in order to truly determine what the reasoning content coding scheme would look like and what would count as pieces of Engineering evidence. In conjunction with the fact that only one Engineering design report was analyzed, our findings are being presented as extremely preliminary work that still needs to be explored further, though they do suggest some possible insights into the differences and similarities between Biology and Engineering as well as the prospect of multidisciplinary applicability of this analytic framework. In order to compare the two disciplines, quantitative statistics were generated of the proportions of different argument units in the particular section of the Engineering design report and compared to the corresponding average proportions found in the relevant Biology laboratory reports. While these were the average proportions of whole laboratory reports when analyzing the Executive Summary and Final Design Description sections, for the Revisions to the Design Solution section only data from laboratory reports that talked about future experimental set-ups was considered. This generated a subset of 14 laboratory reports, where the portions of the report that were proposing future experimental set-ups were extracted and used to generate average quantitative statistics, for example the average proportion of Claims in those future experimental set-up extracts in the subset of laboratory reports generated. We used SPSS Statistics Version 24 to run quantitative analyses and did not run tests for significance as only one Engineering design report was analyzed.

Findings

<u>Reasoning in Executive Summary</u>

Table 12. Percentage of Argument Units in Biology laboratory reports and Executive Summary of the Engineering design report analyzed. 43 Biology laboratory reports were analyzed.

Statistic	Executive Summary	Biology Laboratory Reports
Total Word Count	421	392.8
Total Argument Units	31	31.4
Question (%)	0	1.1
Goal (%)	25.8	0.5

Claim (%)	64.5	62.3
Explanation (%)	9.7	11.5
Evidence (%)	0	24.7

In contrast to Biology laboratory reports, the argument unit Goal occurred much more frequently in this particular Engineering design report's Executive Summary (Table 12). While this could be partially accounted for by the lack of Evidence in the Executive Summary, that still wouldn't necessarily translate into such a large increase in the proportion of Goal argument units. We would propose that this is indicative of an actual disciplinary difference in terms of what are "important results" for the respective fields, especially considering prior research into Engineering disciplinary practices where engineers often have to balance multiple conflicting goals in the workplace (Jonassen et al., 2006). As such, the higher occurrence of the argument unit Goal in an Executive Summary makes sense in terms of portraying the balancing act of Engineering, and the concurrent lack of Evidence as compared to Biology can be explained by the writing prompt's request for "data…[to] remain in the main body of the report".

This can be further supported by looking at the text of the Executive Summary (Figure 11). Goals are used as orienting frames to justify the design decisions that were pursued by the students, with claims and explanations thus demonstrating how the students decided to achieve those goals. This is in contrast to the Biology laboratory reports previously discussed, where students were primarily working with pieces of evidence and claims in order to generate theoretical propositions about Biology, thus providing a possible insight into what the different fields value.

The living wall requested by the Murphy School is intended for use by 3rd and 6th grade students. It is meant to teach the students about sustainable ecosystems by allowing the students the opportunity to grow edible plants during the cold winter months at a small cost. Because students are the primary users interacting with the wall, decisions related to the watering system and structure of the living wall were tailored toward their abilities. In order to ensure that the students are continually engaged with the wall, the design team chose a watering pump that requires manual input. The design choice translates a child's manual efforts into a immediate visual and tangible response, enlightening students about the direct results of their actions.

Figure 11. Passage from Executive Summary of the Engineering design report analyzed. Text in purple was coded as Goal, while text in blue was coded as Explanation, and text in green was coded as Claim.

Reasoning in Final Design Description

Table 13. Percentage of Argument Units in Biology laboratory reports and Final Design Description of the Engineering design report analyzed. 43 Biology laboratory reports were analyzed.

Statistic	Final Design Description	Biology Laboratory Reports
Total Word Count	744	392.8
Total Argument Units	63	31.4
Question (%)	0	1.1
Goal (%)	4.8	0.5
Claim (%)	73.0	62.3
Explanation (%)	17.5	11.5
Evidence (%)	4.8	24.7

While the Final Design Description resembled the composition of the discussion section of the Biology laboratory reports more, with the proportion of Claims and Explanations within a similar value as the Biology laboratory reports (Table 13), there was a surprising lack of Evidence as defined by the coder. While students made claims about their design solutions and what they had done, they did not provide instances of their actual solutions fulfilling those claims or references to photographic evidence; however, this could also be a question of the standards for what counts as Engineering evidence having been set inappropriately by the coder.

For example, looking at a passage from the Final Design Description, only the statement that the holes become "larger the further they are away from the water source" was taken as a piece of evidence for the claim "the holes vary in size", as they provided verification of the statement that there was a variation in size (Figure 12). However, while "Each branch consists of a clear plastic tube" was coded as a Claim (Figure 12), due to no proof being offered of the clarity of the branches or of users considering it clear, it is entirely possible that that is the inappropriate standard to use and that statement should be coded more appropriately as a piece of evidence.

Thus, though the lack of evidence in Final Design Descriptions could be a potentially interesting disciplinary difference, we would like to caveat that due to the methodological issues outlined above, it is extremely important to note that we can only offer this claim tentatively in consideration of the lack of specialist disciplinary knowledge, and that we still need to consult with a disciplinary specialist for what is the appropriate standard to be applied.

To deliver water to the various shelves within the structure, the design team created a system that contains four branches; one for each shelf. Each branch consists of a clear plastic tube, which allows for the students to observe water moving through the system. Ball valves separate each branch, allowing students to water one branch within the structure at a time. Each branch has two sets of two holes that fall over each bucket. To ensure that each plant container receives the same amount of water, the holes vary in size, becoming larger the further they are away from the water source. Finally, Custom, acrylic tube supports were fabricated to ensure that each branch of the watering system remains relatively horizontal, as kinks in the tubing could cause the water to be unevenly dispersed amongst the plant containers.

Figure 12. Passage from Final Design Description of the Engineering design report analyzed. Text in purple was coded as Goal, text in blue was coded as Explanation, text in green was coded as Claim, and text in pink was coded as Evidence.

Reasoning in Revisions to the Design Solution

Table 14. Percentage of Argument Units in Future Experimental Set-ups of Biology laboratory reports and Revisions to the Design Solution of the Engineering design report analyzed. 14 Biology laboratory reports were analyzed.

Statistic	Revisions to the Design Solution	Future Experimental Set-ups
Total Argument Units	20	3.7
Question (%)	0	4.0
Goal (%)	5.0	4.8
Claim (%)	95.0	88.9
Explanation (%)	0	2.4
Evidence (%)	0	0

Similarly to the future experimental set-up extracts parsed from the Biology laboratory reports, the Revisions to the Design Solution section essentially consisted of a string of claims with one additional argument unit (Table 14). This is additionally validated by looking at the text of the section, where the students provide a lot of possible design solutions without giving any evidence or explanations for why their proposed ideas would actually work (Figure 13).

This entails design changes to the planter boxes, which would allow for the integration of a drainage system to prevent plants from being overwatered. This may be as simple as installing reservoirs to the bottoms of the planter boxes, providing space for the water to sit in until it evaporates into the surrounding air. Or as complicated as incorporating more tubing, allowing excess water to actually leave the planter boxes through a drainage system. Implementing a secondary watering system to provide redundancy would also be beneficial for the wall to prevent the plants from being under-watered.

Figure 13. Passage from Revisions to the Design Solution of the Engineering design report analyzed. Text in green was coded as Claim.

This could tentatively suggest that students respond to the idea of future revisions similarly in introductory Biology laboratory reports and final year Engineering design reports in the same way, possibly indicating a shared framing that is activated by the genre of writing out future revisions. The reason why students provide a string of claims, rather than a detailed and well-reasoned argument for their proposed revisions, could be that without a need or opportunity to carry out these revisions, it merely becomes an exercise in "answering the question" (Zeegers & Giles, 1996) as there wouldn't be a chance to turn their proposed claims into reality.

Applicability and Limitations

As a first attempt at exploring multidisciplinary applicability, we would like to suggest that this analytic framework does seem to offer meaningful ways of thinking about different disciplines, and tentatively seems able to uncover the data and signals required to tease apart the differences in how students perform their reasoning different disciplines. However, we also acknowledge the lack of certainty over the full applicability of the argument unit coding scheme, which necessarily colours the preliminary exploratory results we did generate. Thus, we would also propose that more analytic work is needed before we can make stronger claims about the multidisciplinary applicability of this analytic framework, or the differences and similarities between Biology and Engineering written work.

Conclusion

From our preliminary analysis of the data set, we suggest that we have promising results regarding simply reframing Biology laboratory report writing to be explicitly about reasoning and argumentation in order for students to demonstrate more nuanced and novel reasoning, as they provide more Biological conceptual content and less given conceptual content in Year 2 of the study. This would suggest that students do not need the intensive interventions previously recommended in the literature (e.g. Sadler, 2004; McNeill et al., 2006; Kuhn, 2010), and instead that we as instructors and educators need to consider the role our questions and prompts play in possibly causing students to choose not to demonstrate complex reasoning in writing (e.g. Zeegers & Giles, 1996; Peker & Wallace, 2009).

We also tentatively suggest that we have an analytic framework which better accounts for linguistic variation, as it incorporates multiple possible linguistic markers for each factor of reasoning, and which allows us to do both quantitative and qualitative analyses of student reasoning in writing. Moreover, we suggest that there is some evidence that this analytic framework can be applied to other disciplines as well, providing a tool with which texts in different disciplines could be compared. This could offer us a base framework with which to develop analytical instruments that can better characterize student reasoning in all its variation, allowing us to recognize and capture all the ways in which students are reasoning well rather than missing what they're doing because of using a more highly specified framework (e.g. Lawson et al., 2000; Sampson & Clark, 2008).

Bibliography

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching, 49*(1), 68–94. http://doi.org/10.1002/tea.20446

Chomsky, N. (1965). Aspects of the theory of syntax. Oxford, England: M.I.T. Press.

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010) A framework for analysing evidence-based reasoning in science classroom discourse. *Educational Assessment, 15*, 175-196. DOI: 10.1080/10627197.2010.530553

Gasper, B. J., & Gardner, S. M. (2013). Engaging Students in Authentic Microbiology Research in an Introductory Biology Laboratory Course is Correlated with Gains in Student Understanding of the Nature of Authentic Research and Critical Thinking. *Journal of Microbiology & Biology Education*, *14*(1), 25–34. DOI: 10.1128/jmbe.v14i1.460

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, (January), 52–59. Retrieved from http://scitation.aip.org/content/aapt/journal/ajp/68/S1/10.1119/1.19520

Hunt, K. (1965). Grammatical structures written at three grade levels. NCTE Research report No. 3. Champaign, IL, USA: NCTE.

Jonassen, D., Strobel, J., & Lee, C. B. (2006). Everyday problem solving in engineering: Lessons for engineering educators. *Journal of Engineering Education, 95(2), 139-151. doi:10.1002/j.2168-9830.2006.tb00885.x*

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, *86*(3), 314–342. http://doi.org/10.1002/sce.10024

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824. http://doi.org/10.1002/sce.20395

Lawson, A. E., Drake, N., Johnson, J., Kwon, Y. J., & Scarpone, C. (2000). How Good Are Students at Testing Alternative Explanations of Unseen Entities? *The American Biology Teacher*, *62*(4), 249-255. https://doi.org/10.1662/0002-7685(2000)062[0249:HGASAT]2.0.CO;2

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*(2), 153-191. http://dx.doi.org/10.1207/s15327809jls1502_1

Peker, D., & Wallace, C. S. (2009). Characterizing High School Students' Written Explanations in Biology Laboratories. *Research in Science Education*, *41*(2), 169–191. http://doi.org/10.1007/s11165-009-9151-z

Quitadamo, I. J., & Kurtz, M. J. (2007). Learning to Improve: Using Writing to Increase Critical Thinking Performance in General Education Biology. *CBE - Life Sciences Education*, *6*, 140–154. http://doi.org/10.1187/cbe.06

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

Sampson, V. D., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447–472. http://doi.org/10.1002/sce.20276

Sundberg, M. D., & Moncada, G. J. (1994). Creating Effective Investigative Laboratories for Undergraduates. *BioScience*, *44*(10), 698–704. DOI: 10.2307/1312513

Toulmin, S. (1958). The uses of argument. Cambridge, England: Cambridge University Press.

Zeegers, P., & Giles, L. (1996). Essay Writing in Biology: An Example of Effective Student Learning? *Research in Science Education, 26*(4), 437-459. DOI:10.1007/BF02357454