AN INVESTIGATION INTO STUDENTS' PERCEIVED DIFFICULTY VERSUS PERFORMANCE OF SCIENTIFIC TASKS IN GENERAL CHEMISTRY

by

SOPHIA E. NEWTON

(Under the Direction of Norbert J. Pienta)

ABSTRACT

Students' perceived difficulty was compared with performance of scientific tasks typically found in general chemistry courses. These differences were analyzed via direct comparison of a survey (perceived difficulty) administered to participants both directly preceding and following performance of various tasks discussed in survey. The pre/post comparison allowed researchers to determine whether perceived difficulty was affected by task performance as well. A thorough analysis of transcripts of interviews, as well as time on task during performance was also conducted.

INDEX WORDS: General chemistry, chemistry, laboratory, perceived difficulty, task performance, transcript, time on task,

AN INVESTIGATION INTO STUDENTS' PERCEIVED DIFFICULTY VERSUS PERFORMANCE OF SCIENTIFIC TASKS IN GENERAL CHEMISTRY

by

SOPHIA E. NEWTON

BA, Agnes Scott College, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2017

© 2017

Sophia E. Newton

All Rights Reserved

AN INVESTIGATION INTO STUDENTS' PERCEIVED DIFFICULTY VERSUS PERFORMANCE OF SCIENTIFIC TASKS IN GENERAL CHEMISTRY

by

SOPHIA E. NEWTON

Major Professor:	Norbert J. Pienta
Committee:	John Stickney
	Richard Morrison

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

ACKNOWLEDGEMENTS

I would like to acknowledge my major professor, Dr. Pienta, for all his tutelage and mentoring throughout this arduous process. I would like to thank my committee members, Richard Morrison and John Stickney for their guidance. A huge thank you to Lisa Kendhammer for her expertise, suggestions, and ideas. Thank you to Hui (Tom) Tang for always bouncing statistics and R ideas back and forth – this was so very helpful. I'd like to thank my group members, current and former, for their feedback and suggestions during the research process. I'm so very grateful to my family for their support throughout this process: Larry, Jo, Paula, Jade, Lucas, Shaunna, Pat, Fred. Of course I would also like to thank my colleagues at UGA who read drafts of this and remained an extraordinary support system throughout this thesis completion: Patience Sanderson, Lauren Pepi, Molly Atkinson, Ashley Holland, Sahel Mohebbi, Soshawn Blair, and Walter Turner. Finally, the team at Spotify who created the Discover Weekly and Daily Playlist options – I certainly would not have completed this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 GENERAL INTRODUCTION

Studies on students' perceptions of their abilities in laboratory activities, sub-categorized by sex, by class level, by expertise level have been reported. However, a substantial gap in this research exists in comparison of students' perceptions of their abilities with their performance in those laboratory activities. One reason for this deficiency in the literature could be the lack of a consistent, well-vetted rubric for assessing students' performance of laboratory skills in the first place. A well-vetted rubric is one that has been assessed for both its validity (i.e., Is this rubric assessing what it purports to?) and its inter-rater reliability (i.e., Does this rubric yield consistent assessment of participants across reviewers?) [1]. Several rubrics have attempted to categorize the level of inquiry in general chemistry labs[2-8]. However, there is no apparent published rubric for task performance assessment. In the present work, the creation of a rubric for assessment of task performance, the Docktor Robust Assessment on problem-solving in physics students proved most helpfuls[9]. For facile use, rubrics should cover as few dimensions as possible, while still being discriminatory between experts and novices. Any productive assessment should include the criteria being assessed, and the levels of performance[10-11].

Examples of students' perceptions of their experience in the lab appear in the work of Galloway, in which researchers categorized affective learning experiences to assess meaningful use[5]. In its introduction, this paper also explains the importance of constructivism as a meaningful learning framework, and lays out a convincing argument

for the inextricable link between cognitive and affective development. This paper also brings up that in the report on Discipline Based Education Research, the educational community is spending a great deal of time, energy, and resources on laboratory learning for young scientists without the research background to support what, if anything, they are gaining from these laboratory learning experiences[12]. Students' attitudes regarding their chemistry studies, both in the lecture and in the laboratory has been reported[1, 5, 13-20].

In previous studies of students' perceptions of their abilities and their attitudes relating to chemistry and chemistry laboratories, researchers found that students' attitudes towards chemistry were related to their motivation[14]. In 2005, Bauer created a chemistry-specific attitude assessment tool, the Chemistry Self-Concept Inventory (CSCI), a specific application of the Self-Description Questionnaire III (SDQIII), an attitude assessment tool developed in 1984 and repeatedly validated including for various science specific attitudes, intended for use on college-aged adults[16, 21]. The Chemistry Attitudes and Experiences Questionnaire (CAEQ), published in 2002, surveyed students attitudes regarding chemistry specifically on a semantic differential[19]. Semantic differentials include descriptor pairs on opposite sides of an odd numbered line, and allow participants to select the position on the line closest to their perceived fit of the descriptor to the statement given. In 2008, Bauer improved on this design with the Attitude towards the Subject of Chemistry Inventory (ASCI) tool, which was also a seven point semantic differential, but less ambiguous as to the feelings about chemistry sought in measuring with this tool.

> "A prominent theoretical structure for attitude—emerging from factor analysis of results from many contexts and cultures—holds that attitude is composed of components of evaluation (e.g., good–bad, valuable–worthless), potency (e.g., strong–weak, heavy–light), and activity

> (e.g., fast–slow, excitable–calm) *(17, 18)*. In practice, the evaluation component typically explains most of the variance, so the adjectives selected for the ASCI emphasize that aspect."[1]

While there are examples of studies on student attitudes towards chemistry, there are few, if any, assessing student performance of chemistry laboratory tasks, and apparently none comparing an assessment of those students' attitudes about perceived difficulty of and performance of tasks in the chemistry laboratory. This research aims to examine any sex or class differences that may exist amongst general chemistry students' perception of their abilities in the chemistry laboratory, as compared with their performance of those same tasks in the laboratory, and whether those perceptions changed from before performing the tasks to after performing the tasks.

To accomplish the goal of assessing differences in perception versus performance, a survey to determine students' perception of their own abilities was designed. After creating the survey, a rubric to assess students' performance was designed and vetted. Finally, after creation of the survey and the rubric, students were recruited for the study. Subsequently, a survey was designed and administered to students enrolled in the large enrollment laboratory classes a few semesters after the interviews took place. The goal was to see if students' perceptions of skills in those subsequent semesters reflected those of the interview population, and to compare student performance of skills with survey students' responses. The intention was to see how they would apply their previous learning from the laboratory in using skills and equipment in a hypothetical case.

2 INTRODUCTION TO TASK-BY-TASK ANALYSIS

The rubric for each task in this study was designed and created after careful evaluation of the laboratory manuals from which the interview population learned techniques. For this population of general chemistry students, the research team assumed that in-class instruction was the source of these students' primary lab knowledge. .

Description of Interviews

A cohort of 38 students participated in 45-75 minute interviews/observations:, first, they completed a survey that asked about their perceived difficulty of various chemistry laboratory skills; second, they were asked to perform a subset of those skills; and finally, the same survey on perceived difficulty was administered again. The survey portion of the interviews was based on a very similar survey administered to a similar population of students by a research team member, Lisa K Kendhammer, in the Spring of 2015.

After completing the pre-survey, students were provided a brief explanation of the ensuing active interview process, before being asked to begin the task portion of the interviews. Overall there were eight total tasks completed in the interview process; these tasks were performed in the same general order for all interviews. The set order of tasks (rather than randomized task order) was performed with the intention of providing all interviewees with the same experience and not unintentionally giving information in a

different order [22-24]. With the set task order, if priming (i.e., cuing students to an idea or answer) occurred, it should have occurred to all students equally[22]. These tasks were:

1. Reading and recording the volume in a graduated cylinder
2. Performing a 10:1 dilution of an NaCl solution
3. Measuring out and recording the mass of 0.5 g of sugar
4. Transferring 10 mL of solution from one beaker to another
5. Transferring 10 mL of solution from a buret to a beaker
6. Heating a solution
7. Decanting a solution from a solid
8. Performing a titration with a visual indicator

Not all participants completed each the task due to both time constraints and insufficient confidence in their ability to do the task.

Interviews were conducted in a research lab due to the lack of availability of the authentic lab space, and as such were not organized identically to students' familiar lab set up. However, standard lab materials and equipment was made available to the students. Several students asked for "hot hands" (i.e., rubber hot pads for use in laboratories) during the interviews, but because "hot hands" were not available in the fall semester of 2015 in all chemistry 1211 labs, they were not made available for these interviews. Instead, tweezers, beaker tongs, and crucible tongs were available, as they were in all 1211 labs in the fall 2015 semester. A full list of equipment that was made available appears in Appendix B**,**. The laboratory equipment was laid out on the counter, different from general chemistry labs where they are in labelled drawers, but interviewees were given a brief tour of the available equipment before the interview began.

Rubrics Used in Multiple Tasks:

Several rubrics were used in multiple tasks or across all tasks in the interview as a whole.

Those rubrics are included Table 2-1,

Table 2-2, and Table 2-3..

Table 2-1: Selecting Proper Glassware

Exemplary (5)	Selects volumetric flask for dilution. Selects volumetric pipet for transfer 10.00 mL question. Uses funnel to fill buret for titration. Uses stir rod for decanting. Appropriate sized glassware for titration.
Acceptable (4)	Misses one selection from exemplary
Neutral (3)	Misses 2 selections from exemplary
Poor (2)	Misses 3 selections from exemplary
Very Poor (1)	Misses 4 or more selections from exemplary

Table 2-2: Recording Data

Exemplary (5)	Records volume in graduated cylinder task 1 Records volume of NaCl transferred in task 2 Records mass in task 3 Records initial & final buret readings in task 5 Subtracts initial & final buret readings in task 5 to confirm amount transferred Records initial & final temperatures in task 6 Records amount of NH_3 transferred to receptacle in task 8 Records initial & final buret readings of HCl in task 8
Acceptable (4)	Misses 1 – 2 from exemplary
Neutral (3)	Misses 3 – 4 from exemplary
Poor (2)	Misses 5 – 6 from exemplary
Very Poor (1)	Misses 7 or more from exemplary

Table 2-3: Recording Observations

Exemplary (5)	Records identity of solid in task 3 Records approximate time* in task 6 Records initial color of indicator in task 8 Records color on addition to reactant flask/beaker in task 8 Records color at end point of titration in task 8 Records color past end point in task 8 (if applicable)
Acceptable (4)	Misses task 3 or 6 recordings, but records at least one color in task 8
Neutral (3)	Misses task 3 & 6 recordings but records at least one color in task 8
Poor (2)	Misses task 3 or 6 recordings and does not record observations in task 8
Very Poor (1)	Does not record any observations throughout 8 task interview

It is important to note that these rubrics were created based on the contents of the laboratory manuals created by Dr. Daphne Norton for exclusive use by the University of Georgia students in the general chemistry sequence. Since these courses are at an introductory, 1000 level series, students are not expected to have significant background knowledge before completing/enrolling in the series. Because this is the case, the bulk of students' laboratory knowledge is expected to have come from these manuals and any supplemental information available to them via instruction by their teaching assistants. However, their knowledge and skills could also come from their previous laboratory experiences. As a result, the laboratory manuals were chosen as the authoritative source even though laboratory manuals from other sources and even the experiences of this researcher suggests alternatives. As such, these rubrics were created and students were subsequently graded on the primary and pre-eminent source matter.

The use of rubrics is intended to minimize grader bias, and in the present study, to assure consistency across all the activities. Because of the shortage of literature precedents, best efforts were made to create robust rubrics. All comparisons of survey data to these grades are subjective and subject to interpretation, a potential shortcoming

in the research plan. One way to address this in a future study would be to use a panel of graders, but in the current study with 38 interviews, the data represents the best efforts of a single researcher. A major accomplishment of this study is that it represents a systematic approach to assessing student performance of laboratory tasks and to creating rubrics that are part of such assessment. As such and as an application of this work these rubrics could be incorporated into future students' grades within the courses, and into further vetting of the rubrics themselves. Inclusion of a rubric into their laboratory manual by which students could be graded could be an extraordinary tool for students, allowing them to study and to become very familiar with the expectations of them. Towards that end, another tool could be inclusion of illustrated task performance into the lab manual, or videos of task performance to be viewed before performance of tasks in lab for a grade.

This research project is a set of suggestions for grading and for understanding students' performance quantitatively as well as qualitatively, based on the information for which they are responsible upon completion of the series.

Ultimately these rubrics are an appropriate measure by which to compare interview performance, both to one another (across interviews) and to their perceptions of their performance and abilities via their surveys (within participants).

Student Demographic Breakdown

The 38 person interview population comprised26 female and 12 male participants. At the time of the interviews, the course enrollment comprised xx% females and yy% males. Thus, when normalized for comparison, there was not a significant difference in

female/male ratio of course enrollment. This number was also compared to the percent sex enrollment at the University of Georgia as a whole, and within its Franklin College of Arts and Sciences, and again, no significant difference was found between sex distributions enrolled.

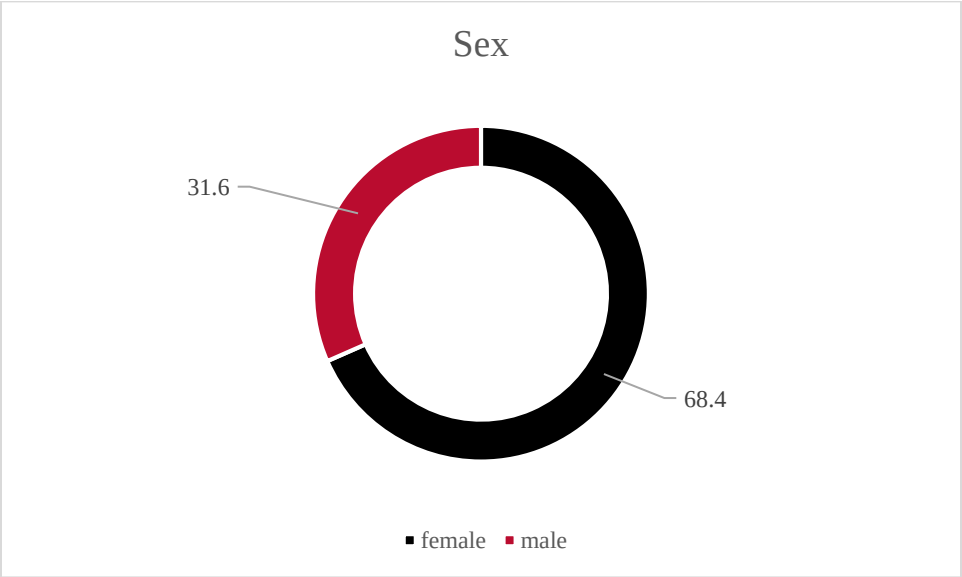

Figure 2-1: Sex breakdown of interview population

For the university as a whole, during the fall semester of 2015, there were 11,864 male students enrolled and 15,568 female students enrolled, or a 43% male/57% female student body population. When broken down by class, there were 2119 male and 3173 female first year students, or a 40% male/60% female first year student body. Since the interview population was 32% male / 68% female, a two sample t-test between percentages was run for the male and female populations, both as undergraduates as a whole and first year undergraduates to the interview population, and no significant differences were found between populations.

This breakdown is illustrated in Figure 2-2: Sex breakdown of interview population (inner) vs undergraduate population (outer) Figure 2-2, on page 10. There was no significant difference found between the populations,

t (15592) = 1.224, p = .2213.

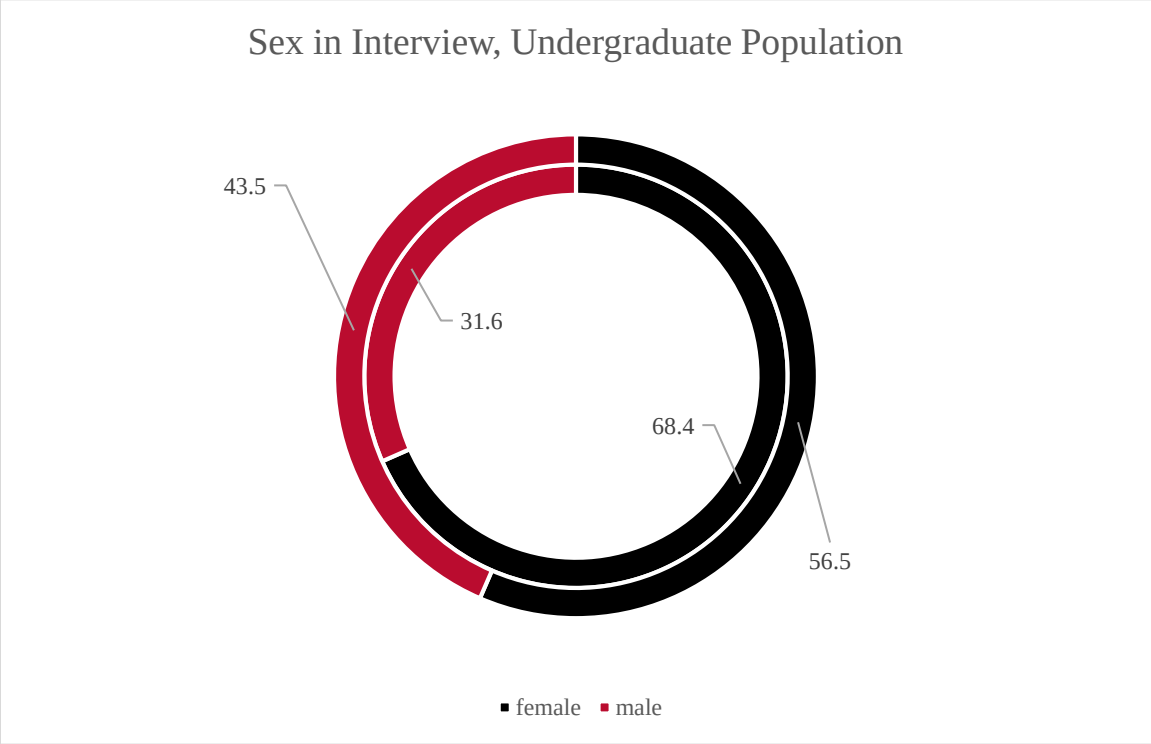

Figure 2-2: Sex breakdown of interview population (inner) vs undergraduate population (outer)

There were 20 1211 and 18 1212 students in the interview, which roughly compared with the combined 1211/1212 populations of the Fall 2015 and Spring 2016 populations.

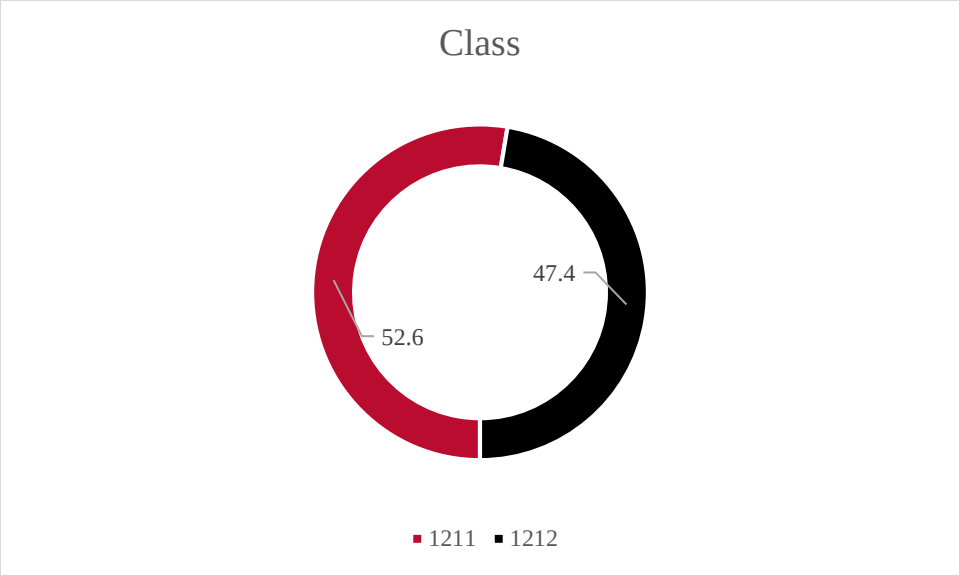

Figure 2-3: Class breakdown of interview population

Class breakdown of the interview population is illustrated above, in Figure 2-3 – there were 53% 1211 students, 47% 1212 students interviewed.

Analyses Performed on Each Task

For each task, applicable pre-and post-interview survey items, as well as rubric items were compared for meaningful differences via both ANOVA and paired t-tests. The results of the pre- and post-interview surveys were directly compared via t-tests. Ultimately, one of the goals of this study was to compare perceived difficulty both before and after performing tasks, to task performance itself. Few significant differences between pre- and post-interview survey scores were observed, but significant differences were observed between their task performance scores compared with their survey scores. This is a multi-faceted problem: these rubrics were subjective based on scorer, the survey items were subject to participant scale, and these scales are not identical. It is difficult to compare (even on a comparably numbered scale) across participants for difficulty level,

and more difficult still to reconcile those differences with a less-subjective score on the task.

.

Significant differences in task performance (via the rubrics) and perceived difficulty (across and within participants) appeared on several tasks. Although these differences could be ascribed to the effectiveness and match of the rubric to students' actual behaviors, it is much more likely that the students' perceptions of task difficulty was significantly different than their skills at performing the task. One way to test this difference would be to survey students on which glassware or equipment they think they would use, determining whether they would use the correct glassware in each task and therefore be penalized under this rubric. Another way to test this hypothesis would be to modify the rubric, and grade students according to the new rubric as well, and see whether there was still a significant difference in perceived difficulty and task performance with the modified rubric. Finally, a form of vetting that was considered (but ultimately not performed due to time constraints) for this research project was an inter-rater reliability study. By having multiple raters grade each interview task, and revisiting or re-evaluating any interviews scored such that interviews significantly disagreed with one another on student performance of that task, the reliability of the scores themselves could be tested.

Interviews were analyzed for time on each task as well, and compared across sex and class sub-populations. Time on task was measured in seconds and recorded as a part of the transcription process, which was carried out entirely by the primary researcher.

Time on task began with reading of instructions and was complete when interviewees indicated that they had completed the task, and at the researcher's discretion when no indication was made explicitly. For some tasks (e.g., dilution and titration), there appeared to be a separate planning versus action phase for some participants. Without a means to discern this difference for all students and because some participants would go through several iterations of planning/action cycles, only full time on task was analyzed.

Interviews and their transcripts were also analyzed via their transcripts for language to elucidate participants' thought process by speech density (# words / time on task) per task and throughout interview as a whole. Most frequent terms were collected for each task (as well as across whole interviews) and displayed both in histograms and word clouds to expedite visualization of the language students were using in their interviews.

While none of these analyses alone could describe the relationship between perceived difficulty and task performance, taking these together paints a picture of this relationship.

3 ANALYSIS OF TASK 1: READING A GRADUATED CYLINDER

Introduction to the Task

The first task in each interview, after completing IRB and pre-survey paperwork, was to read and record the volume in a 50-mL graduated cylinder. Since this graduated cylinder was marked to the units place, according to their lab manual, they should've recorded to a precision of one decimal place, as well as recorded the unit of measure, mL.

The researcher anticipated students recording a different number of digits than was required, neglecting to record units for their measurements, and reading the volume from other than eye level.

Common Interview Behaviors

Out of 38 students interviewed, 7 (18.4%) recorded a different number of significant figures than required, and 5 did not record a unit on their measurement. Not one student read the volume of the graduated cylinder from elsewhere than eye level, indicating that the importance of reading the meniscus of a solution at eye level has been effectively taught to all general chemistry students in the program. Interestingly, two students (5.26% of those interviewed) did not record a volume or unit at all, even though this was explicitly a part of the task given to them. This could have reading comprehension implications, or could be chalked up to general nervousness with being recorded and interviewed in the first place. Ten students, more than one quarter of the interview population, used their fingers or a pen as a pointer, counting up to the meniscus

level from a nearby “landmark” on the cylinder. It is unclear whether this was a technique taught to them by their instructors, or if they naturally are inclined to count up from landmarks. Since this techniques’ prevalence did not become apparent until the data analysis phase of this research, students were not asked follow-up questions about this behavior – this would need to be a planned question for future research if the team was interested in why that occurred.

Time on Task

Students’ time on task for the first task, reading from a graduated cylinder, is displayed below, in Figure 3-1.

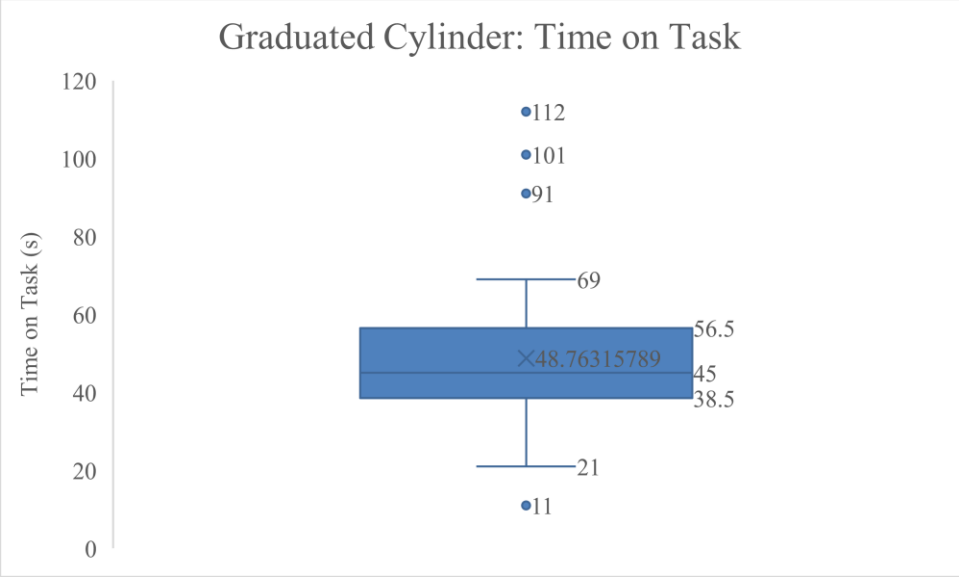

Figure 3-1: Box and Whisker Plot of Time on Task for Task 1

Students spent an average of 49 seconds on the first task (read and record the volume in this graduated cylinder). The fastest participant finished in 11 seconds, while

the slowest spent 112 seconds on the task. Only four students' response times fell outside the standard range covered by a box and whisker plot.

From 020203:

> "Okay so, I'm going to get down on eye level and I'm going to look at where the little um meniscus is, the little dippy thing, and it looks like it's exactly on the 30 1, 2, 3, 32 line, it might even be a little bit above that so I'm going to do 32.1 mL."

Within each task, there were several instances of participants who were chatty, as well as of participants who said very little. The above quote was from a particularly wordy participant, they said nearly 5000 words in the interview (average: +/-). While this participant did use relaxed language including several filler words, they also used a naming word (meniscus) and it was clear that they were making efforts to make their thoughts and actions well understood by anyone watching this footage at a later date. Overall, participants took care to read the whole task aloud and narrate their actions (and sometimes the thoughts behind those actions) for this task. The average participant said 50 words for this first task, and took 49 +/- 20 seconds to do so.

There was one participant who (despite being instructed to read the task aloud, then narrate their actions throughout the task, same as everyone else) did not say a single word throughout the first task. This participant took 40 seconds to complete task one, which is a little less than the group average of 49 seconds, but well within the acceptable range (this was not an outlier performance of task 1, in terms of time on task – just in terms of words used).

Survey Skills Tested in Task 1

Task 1 tested survey skills of reading a graduated cylinder, knowing what data to record, and knowing what observations to record. The rubric for this task was as follows in Table 3-1

Table 3-1: Rubric for Task 1: "Read and record the volume in this graduated cylinder."

Exemplary (5)	Reads at eye level. Uses correct number of significant figures Records value with units Value is within +/- 0.2 mL of researcher's recorded value

Acceptable (4)	Within +/- 0.3 mL
	One of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Neutral (3)	Within +/- 0.5 mL
	Two of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Poor (2)	Outside +/- 0.5 mL
	Three of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Very Poor (1)	Outside +/- 0.5 mL
	Four or more of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Participants were expected to leave the graduated cylinder on the counter, bend or kneel to eye level with the meniscus, and read the volume in the graduated cylinder. After reading the volume, participants were expected to record the volume to one decimal place and record the unit, since this was a 50.0 mL graduated cylinder with 1 mL precision. According to the Norton lab manual, graduated glassware should always be read to one more degree of accuracy than can be read explicitly from the glassware, ie for a 1 mL precision graduated cylinder, read to the tenths place, and for a 0.1 mL graduated cylinder read to the hundredths place. Significant figures were a focus for a large chunk of points for the first several labs in the 1211 course, so it was expected that students were able to correctly gauge significant figures in a short amount of time.

Survey and Scoring Data

In the pre-interview survey, reading and using a graduated cylinder were split into two separate questions to determine their difficulty. The task of reading and recording the volume does not explicitly split the two, but does only require reading, not use of, a graduated cylinder. In those pre-interview responses, 34 out of 38 students surveyed perceived that reading a graduated cylinder was easy or very easy, while 37 out of 38 students surveyed perceived that using a graduated cylinder was easy or very easy. These response distributions can be viewed in Figure 3-2 & Figure 3-3, below.

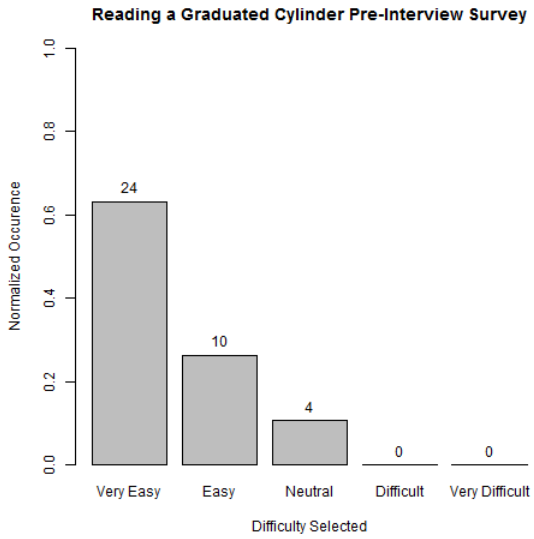

Figure 3-2: Responses from Pre-Interview Survey to Reading a Graduated Cylinder.

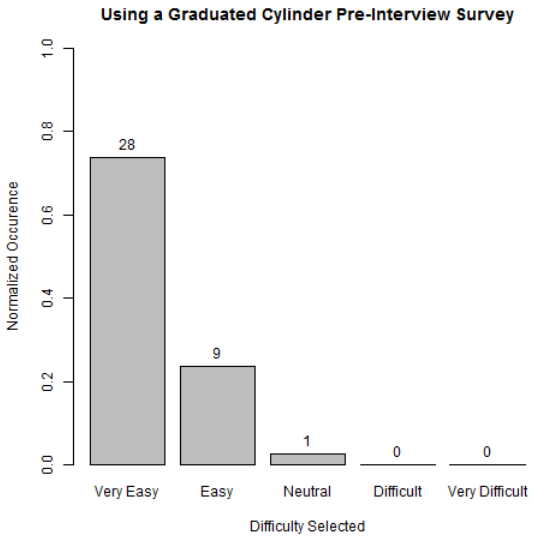

Figure 3-3: Responses to Pre-Interview Survey Using a Graduated Cylinder

During the interview's task performance portion, 32 students received a score of 4 or 5 on the reading a graduated cylinder portion, and 22 received a 4 or 5 on the use of a graduated cylinder portion. While this could, again, be a point of contention amongst other researchers, the explicit wording of task 1 did not constitute use of a graduated cylinder to this researcher, but merely reading a graduated cylinder. Use of a graduated cylinder would mean using a graduated cylinder to transfer an aliquot of solution from one piece of glassware to another, or using that graduated cylinder to measure (but not transfer) a specific amount of solution. This is why only 30 of the 38 students interviewed have a grade for Using a Graduated Cylinder – not all students interviewed completed that task.

Students were not asked during the interviews whether reading a graduated cylinder was perceived to be a part of using a graduated cylinder. It was considered to be

a part of using a graduated cylinder to the interview team, given one cannot properly use a graduated cylinder fully without also reading a graduated cylinder. For this reason, if an interviewee did not read the graduated cylinder properly during the interview task performance, they were marked down for use of a graduated cylinder as well. Dispersion of students' scores on reading a graduated cylinder are displayed below, in Figure 3-4 & Figure 3-5.

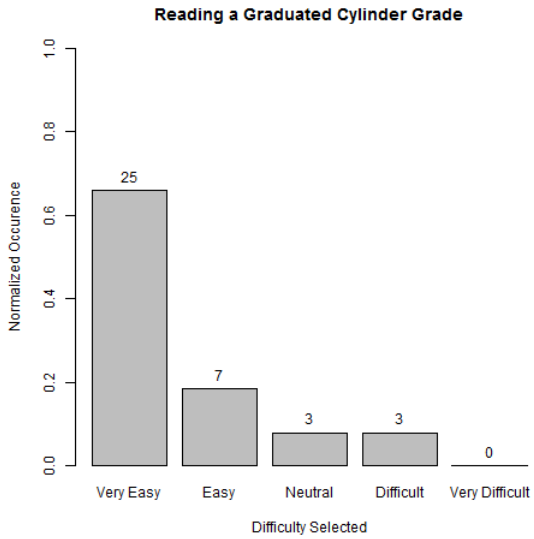

Figure 3-4: Scores to Task 1, Reading a Graduated Cylinder

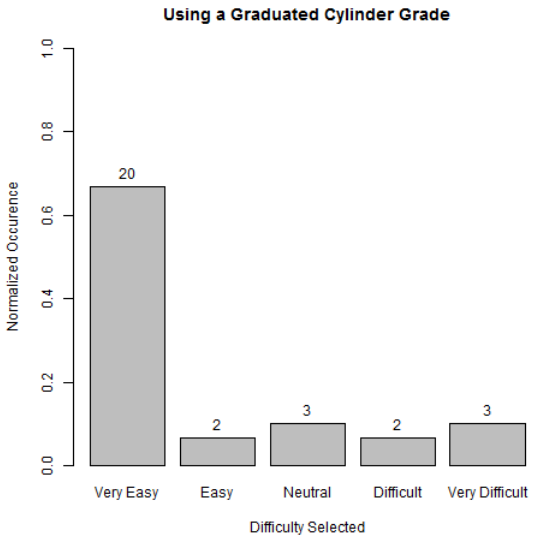

Figure 3-5: Scores to Task 1, Using a Graduated Cylinder

The most common error amongst interviewees for the first task was use of the wrong number of significant figures (7/38, 18%), followed by not including units in recorded volume (5/38, 13%). Of those interviewed, 3 participants did not record a volume (or a unit) despite being given the same task, which explicitly read to record the volume in the graduated cylinder. Something that wasn't an error, but was very common amongst the interviews was use of writing utensil or finger as a pointer while counting

the marks on the cylinder (10/38, 26%). It is possible this behavior could be a discriminatory behavior between experts and novices in chemistry laboratory settings[25]. Out of 38 interviewees, none read the cylinder from anywhere other than the bottom of the meniscus. The word meniscus was mentioned 8 times in this task as a whole.

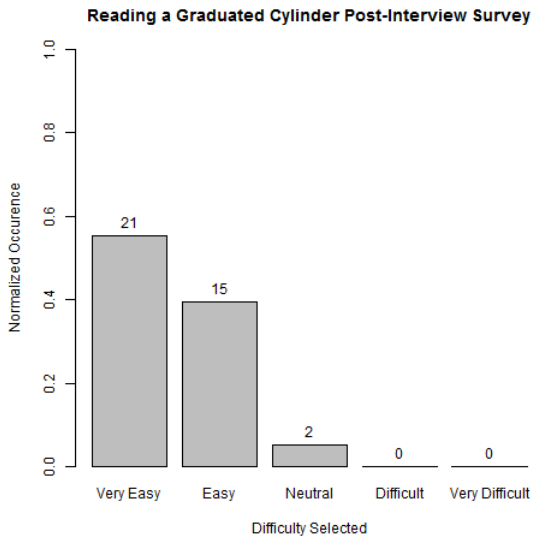

Figure 3-6: Responses to Post-Interview Survey Reading a Graduated Cylinder

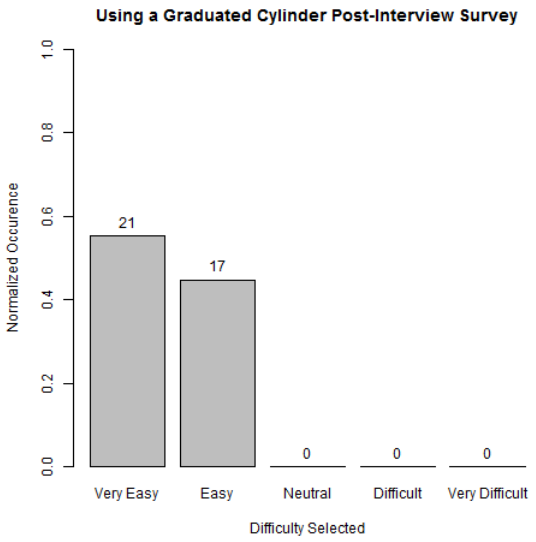

Figure 3-7: Responses to Post-Interview Survey Using a Graduated Cylinder

In the post interview survey (response distributions in Figure 3-6, Figure 3-7, above), 33 students responded that reading a graduated cylinder was easy or very easy, while 35 responded that using a graduated cylinder was easy or very easy. This does represent a change from the pre-interview survey in which 5 total students described reading and using a graduated cylinder as neutral. Neither of these represents a significant difference in pre/post interview scores to these tasks: participants did not perceive a difference in task difficulty for reading a graduated cylinder before (M = 4.51, SE = 0.12), or after (M = 4.54, SE = 0.10) performing the interview tasks, t (34) = -0.298, p = .768. Participants also did not perceive a significant difference in task difficulty for

reading a graduated cylinder, t (34) = 2.026, p = .051, before (M = 4.71, SE = 0.09) or after (M = 4.51, SE = 0.09) performing the interview tasks.

Analysis of Task 1 Transcripts

A histogram of the ten most frequently occurring terms in the first task is found on page 22 in Figure 3-8. These ten terms are, in alphabetical order, and, cylinder, graduated, its, read, record, the, this, volume and you. It was interesting to see "and" and "the" appearing in this list, since special care to remove English stopwords from the analysis for this graph, and according to a list of the stopwords specifically in this mining package, "and" and "the" are stopwords[26].

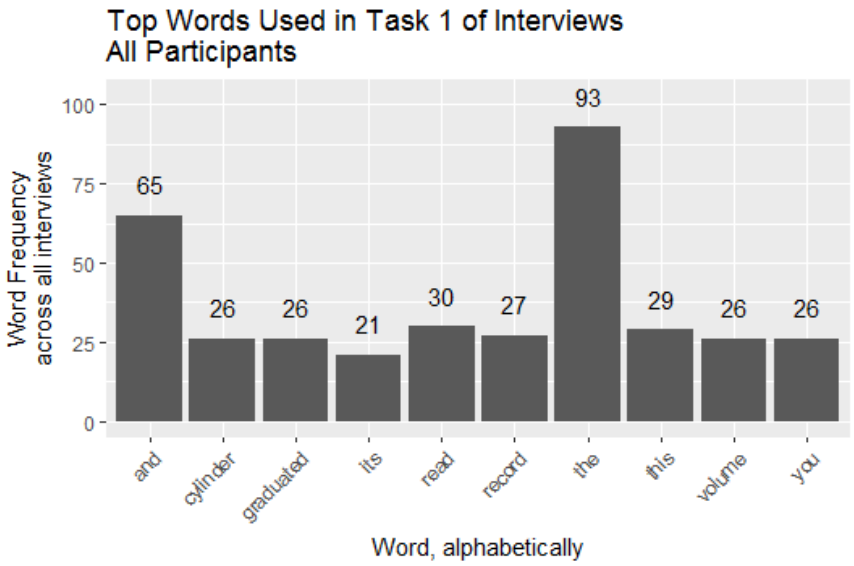

Figure 3-8: Histogram of most commonly used terms in Task 1

During the first task, the most frequently word used by students was "the", with 93 instances. Given that there were 38 interviews, and that students were explicitly instructed to read the task aloud before performing the task, it is interesting that there are

only 26 instances each of “graduated”, “cylinder” and “volume”, with 27 of “record” and 30 of “read” when the task read “Please read and record the volume of this graduated cylinder.” – but it turned out this was because the prompt was removed from transcripts before analyzing. Though there are less than 30 instances of words within the prompts, this is not cause for concern, since one reading of the prompt per participant was removed from transcripts. Therefore, the 27 occurrences of the word “record” meant that participants repeated the word “record” 27 times outside of their initial reading of the prompt.The most commonly used 20 terms during the first task are also displayed in a word cloud, below in Figure 3-9.

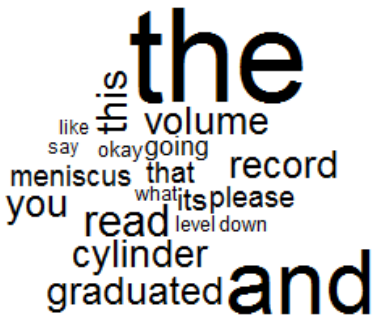

Figure 3-9: Word Cloud of 20 commonly used terms in Task 1

While stopwords consistently appear, scientific words like volume, level, read, graduated cylinder, and meniscus also appear prevalently.

Overall, there was not an overwhelming amount of new information discovered with the first task of these interviews: this task was intended to collect real data, while easing interview participants into the interview process and without overwhelming them with

too many new experiences at once. Participants overwhelmingly fared very well, receiving 25 scores of 5 (exemplary) on reading a graduated cylinder, and 20 scores of 5 on using a graduated cylinder. This matched with their perceived difficulty of both items, which were also overwhelmingly ranked as very easy (5) or easy (4). There was not a significant difference between their perceived difficulty of these tasks and their performances of them. Many students mentioned the meniscus specifically when talking through their performance of the task, indicating an understanding of how to appropriately read a graduated cylinder.

4 ANALYSIS OF TASK 2: DILUTION OF A SOLUTION

Introduction to the Task

For the second task of the interview, students were asked to perform a 10:1 dilution of a sodium chloride solution. While the researcher intended to see students' problem solving skills, application of the dilution equation, $M_1V_1 = M_2V_2$, and performance of a dilution in a volumetric flask with a volumetric pipet as the transfer vessel for their concentrated sodium chloride solution, this was not, overall, what happened.

Table 4-1: Rubric for Use of Volumetric Flask

Exemplary (5)	• uses volumetric pipet to place 20.00 mL NaCl into flask initially. • Dilutes with H_2O to line. • Caps and inverts after adding some but not all of water. • Slows pouring at narrowing of neck. • Uses transfer pip to add water dropwise to line. • Caps and inverts again to ensure proper mixing. • Uses funnel.
Acceptable (4)	• Does not use funnel to pour and/or does not use transfer pipet for last bit to line • Still doesn't fill past line OR • Uses large graduated cylinder (loses decimal place) to put 20 mL but otherwise correct
Neutral (3)	Adds 20.00 mL to flask initially, generally uses properly but either • Does not cap + invert • Fills (small amount) past line
Poor (2)	• fills past line and does not invert. • Does not use funnel, • does not use transfer pipet.
Very Poor (1)	• Doesn't use line. • Still uses 100 mL dilution.

Table 4-2: Rubric for Performance of Task 2, Dilution

Exemplary (5)	• Use volumetric flask without prompt • Perform calculations correctly • Use volumetric pipet twice • Fill to neck with funnel/pour, then use smaller pour/transfer pipet to finish fill to etch
Acceptable (4)	
Neutral (3)	• Use graduated cylinder/other glassware but dilution is perfect
Poor (2)	
Very Poor (1)	• Does not complete dilution • Uses beakers as exact measures of volume • Uses volumetric flask (200 mL) to perform 10/100 mL dilution

Discussion of Time on Task Information

For some students, there was a very clearly defined planning period followed by performance of that planning to complete the dilution. For others, they were completely stymied by the vagueness of the task and either asked to move on or took the researcher up on the opportunity to return to this task at a later time. For others still, there was not a clear division between planning and performance stages. These students seemed to jump right into performance of the task, to mixed results.

Of the students who had a clear division between planning and performance stages (6), the average planning time was 170 seconds, while average time spent performing the dilution was 230 seconds. The range of planning time was 70 - 205 seconds, and range of performance was 75 - 400 seconds. Median planning time was 200 seconds and median performance time was 278 seconds.

For the students who struggled to decide how to approach the task, the average time on struggle was 25 seconds, ranging from 5 – 75 seconds with median 40 seconds. Of those students who moved on, 4 actually had time at the end of the interview to return to the task, 2 of them did attempt the task again. Of the students who did attempt the task

again, 1 completed the task with a dilution (regardless of correctness) while 1 still did not complete a dilution at all.

Discussion of Common Interview Behaviors

Overall, two students completed the dilution in a volumetric flask correctly without prompts. Four used a volumetric flask, but incorrectly. A full 19 (50%) completed the dilution in a graduated cylinder or other glassware than a volumetric flask. Of those 19, 6 were graduated cylinder diluters. Ten were prompted with a graduated cylinder, of those 10, seven completed the dilution with the volumetric flask (in varying degrees of correctness). Across the whole interview population, only two overfilled the volumetric flask (of the # who used a volumetric flask in total) by going past the etched line on the neck of the flask. Four used a transfer pipet to fill to the etched line, to avoid going past the line. Four used a funnel to transfer water into the flask, and of those four, (#) kept the edge of the funnel off the edge of the flask to avoid vacuum creation. (#) overfilled because of the funnel's vacuum creation.

Out of 38 interview subjects, (#) of whom completed calculations for the dilution, 8 (21% of population total, #% of calculation completers) performed the calculation for their dilution incorrectly. Most frequently within this group, students stated that they needed an 11:1 dilution (ie 10 mL NaCl and 100 mL distilled water, rather than 10 mL NaCl and 100 mL total diluted solution). This is a common misapplication of $M_1V_1 = M_2V_2$.

Discussion of Interview and Survey Statistics

Although 32 total students perceived the task of using a volumetric flask during the pre-interview survey as being easy or very easy, and 28 perceived selecting the proper glassware to carry out a task as being easy or very easy, only 6 students actually used a volumetric flask to perform a dilution, without prompting from the interviewer. With regard to scientific skills used in this task, knowing what data to record and creating your own procedure, interviewees collectively thought that recording the correct data was easy overall, (M = 3.71, SE = 0.18, 24 thought it was easy or very easy), while creating your own procedure was perceived as neutral overall, (M = 2.59, SE = 0.20, 5 thought it was easy or very easy). Selecting glassware was perceived as very easy overall, (M = 4.02, SE = 0.23, 28 thought it was easy or very easy), and using a volumetric flask was perceived as very easy overall, (M = 4.17, SE = 0.70, 32 thought it was easy or very easy).

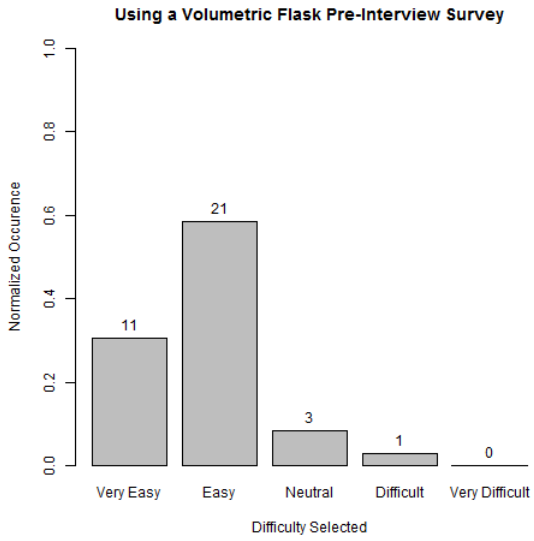

Figure 4-1: Using a volumetric flask, pre-interview survey

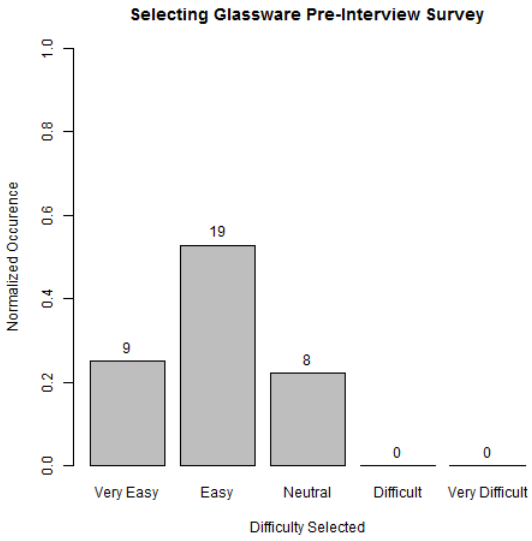

Figure 4-2: Selecting glassware, pre-interview survey

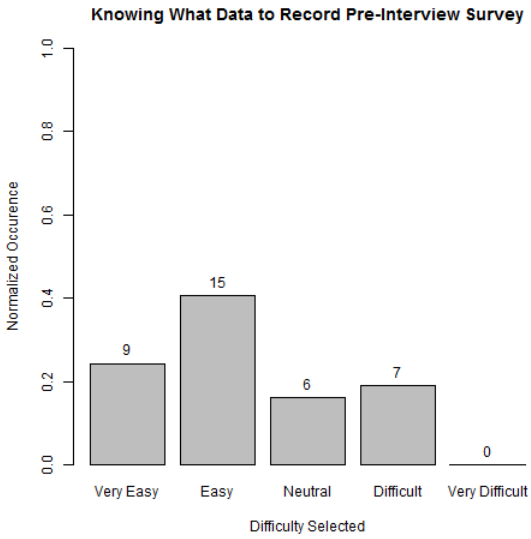

Figure 4-3: Recording data, pre-interview survey

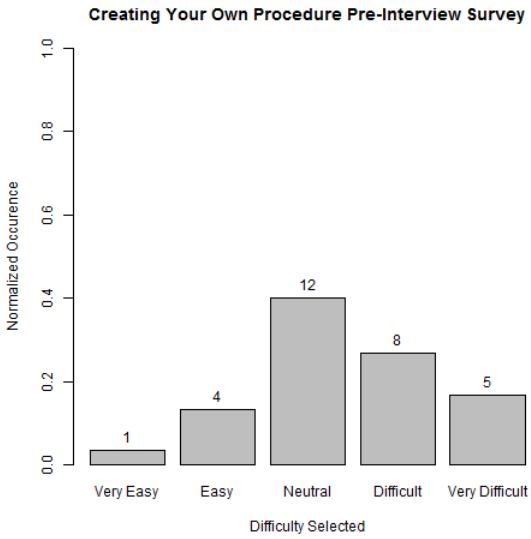

Figure 4-4: Creating a procedure, pre-interview survey

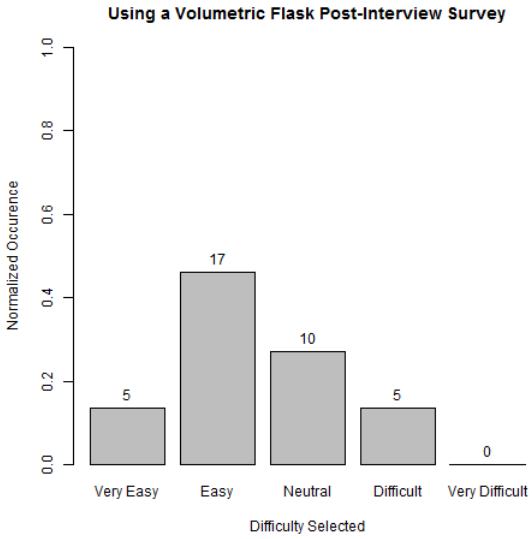

Figure 4-5: Using a volumetric flask, post-interview survey

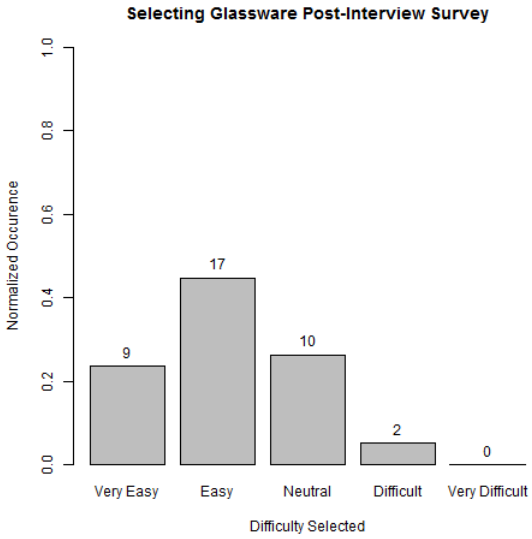

Figure 4-6: Selecting glassware, post-interview survey

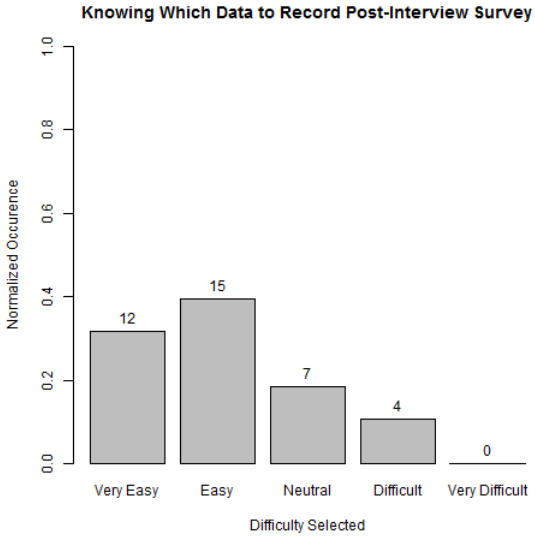

Figure 4-7: Recording data, post-interview survey

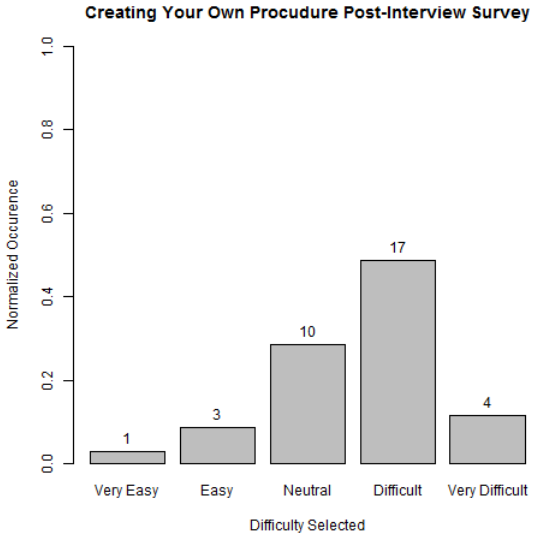

Figure 4-8: Creating a procedure, post-interview survey

On the post interview survey, students perceived that selection of the correct glassware to perform a task had gotten more difficult, though not significantly so: (M = 3.94, SE = 0.14, 25 selected easy or very easy), while use of a volumetric flask had also become more difficult, (M = 3.56, SE = 0.16, 19 students selected easy or very easy). This difference, - 0.625, was significant, $t(31) = 3.507$, $p = .001$, with a large-sized effect $d = 1.20$. This means that students perceived using a volumetric flask as being substantially harder after completing a dilution task, and this was a statistically significant effect. Procedure creation was now perceived as difficult, (M = 2.14, SE = 0.36, 4 thought it was easy or very easy). Those 4 were not included in the initial 5 who thought this was easy or very easy during the pre-interview survey.

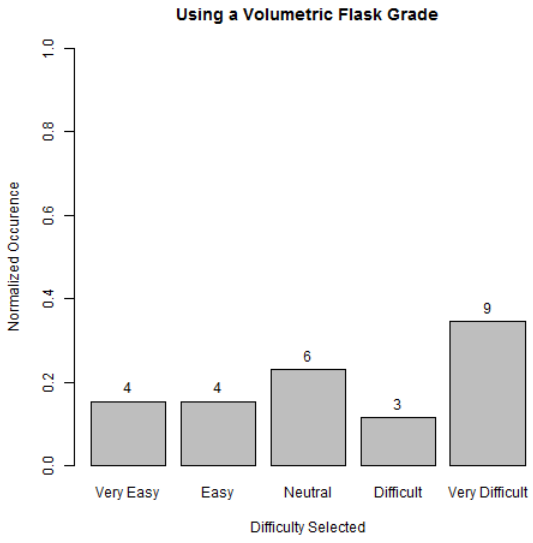

Figure 4-9: Interview Score, Using a Volumetric Flask

Score distribution according to the rubric in Table 4-1, of interview subjects who completed a dilution with a volumetric flask (26 total students) is shown in Figure 4-9. 9 students scored a 1, or very difficult, 8 total students scored a 4 or 5 – so those would

correlate to easy or very easy scores on the pre- and post- interview surveys. This distribution of scores is both an indictment of students understanding of dilutions and the researcher's creation of a rubric.

On realizing that students would likely complete their dilutions in graduated cylinders or other glassware, the researcher decided to ask interviewees their intended use of this dilution, their answers were mostly variations on "this is the final use, is to make the diluted solution" – which is not in fact the final use of these solutions. The disconnect between why they are being asked to perform a task, and their being asked to do the task, could explain the creation of such small amounts of diluted solutions. If they are not thinking forward to utility, it could make perfect sense to make only 10 mL of a diluted solution. Critical thinking and anticipation of utility do not seem to be key priorities of these students, merely getting the scores they desire on their assignments. This is a theme throughout all of the tasks in these interviews.

Discussion of Transcripts

The eleven most commonly used words in the second task of these interviews, dilution of a solution, are displayed in the histogram in Figure 4-10, above. These words are, alphabetically: going, just, know, like, molar, NaCl, need, okay, solution, use, and water. These are categorized as scientific terms, which makes sense given the verbal nature of this task.

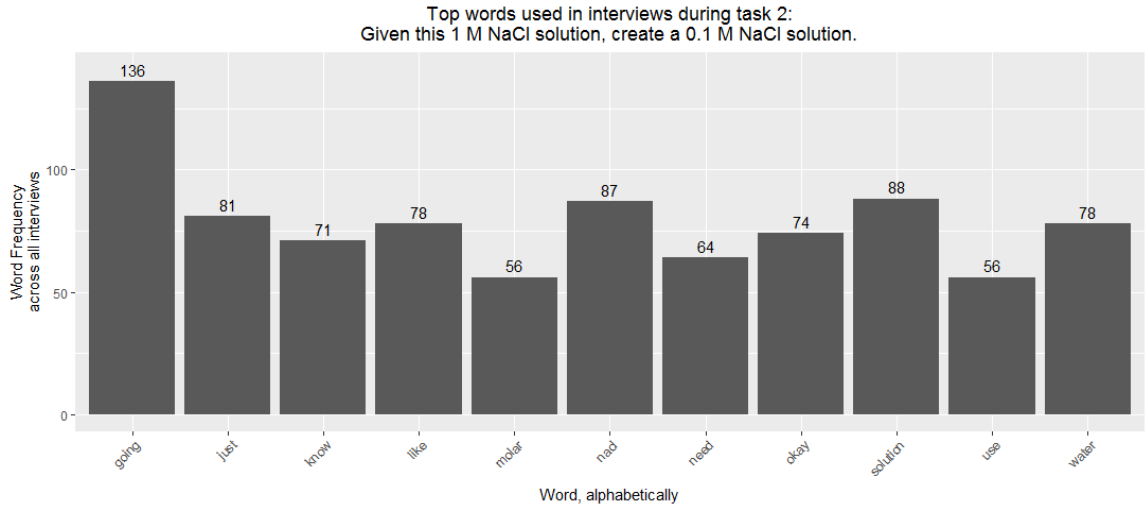

Figure 4-10: Histogram of 11 most frequently occurring terms in Task 2

While users often asked to move on and return to this task, rather than finishing it in order, users did often talk through possible action plans while making the decision whether to move on or continue to struggle with this task – which explains the heavy occurrence of "just", "know", "like", and "okay". Students used these words often with their internal (externalized) debate of what to do next:

From 012803:

> "Okay so uh, I want to work it out to where I have an initial and final volume, and with uh, 100 mL of solution, and that's going to be my ending volume, and I know that my ending molarity is going to be a .1 M, and I'm going to set up like a C1V1 to my known molarity and then how much mL I'm going to use of that's going to be x, so um…"

This participant was partially stalling, attempting to decide their next step while still working on the problem, but approaching it in a scientific way. They have decided what they know, what they need to know, and how they're going to get what they need to know from what they already do. Ultimately, this is what we want students to do to complete effective problem solving when they don't know how to proceed.

On the other hand, participant 111102:

> "So I'm pretty sure that is the correct dilution, I just don't know if I have .1 molar."

After having completed their task. This participant had just completed a dilution of 10 mL NaCl, 100 mL water in 250 mL beaker. This indicated that the participant was (incorrectly) confident in the correctness of the task which they had just performed, but not confident in the calculations which had led to performing the task the way they had.

In Figure 4-11, below, a wordcloud of the 50 most commonly occurring terms in the performance of the dilution task. This and all task specific wordclouds were not stemmed; that is to say that words like mole, moles, molar, and molarity were not consolidated to be separate occurrences of the same stem word. The researcher made the decision to keep these wordclouds unstemmed so that there was some indication of frequency of use for molarity versus moles. In the case of this task specifically, it turns out that this was a trivial difference, with molarity occurring 42 times, molar occurring 56 times, moles occurring 33 times, and mole occurring 43 times. Over the span of these interviews, there were 602 distinct terms, with 3248 total occurrences. This means that these configurations of mole accounted for about 5% of total words said during this task.

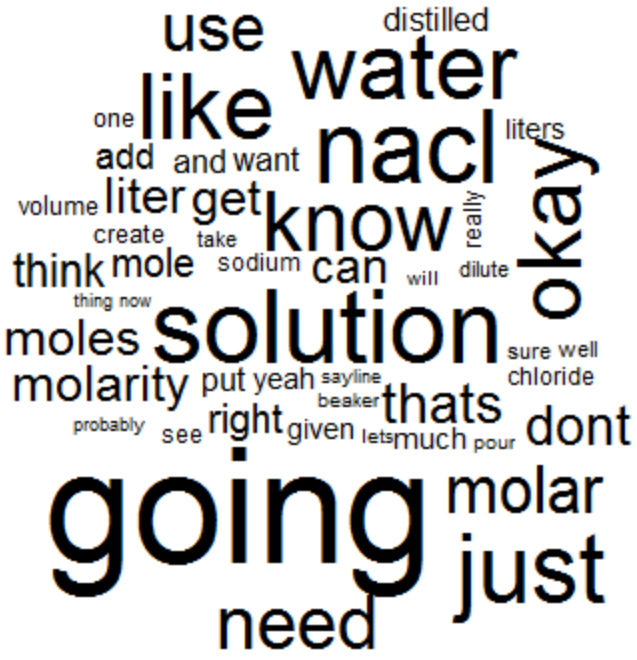

Figure 4-11: Wordcloud of 50 most commonly occurring terms in Task 2

Conclusions from task 2

Overall, while task 2 (dilution) was offered to all 38 participants, 12 completed the task. Of those, 4 completed the task with the expected glassware, prompted or unprompted. Several students performed the dilution but with trivial amounts of solution, which indicated that they were not considering that further actions may be performed with this diluted solution. Not one single participant asked what this solution would be used for, so this researcher hesitates to speculate on what they anticipated doing with the new solution. Overwhelmingly, although many students elected to return to the task at the end of the interview or to skip it entirely, students' language reflected a combination of scientific planning/problem solving and stalling while performing that planning/problem solving. While performance of this task in the way the researcher intended was difficult for participants overall, students still felt confident in their performance after the

interview was complete. Implications for further studies on the match of perceived difficulty versus performance of dilutions include a need to assess fitness of a dilution for purposes intended, and that there needs to be a vein of research on what students expect (if anything) to do with the solution they have created, or if creation of the solution itself is the end goal. Although students do not seem to hold the same viewpoint, many of their approaches to this task were akin to creating a few tablespoons of broth for a soup recipe – while they have technically done what they have been asked (created a broth) they haven't always made enough to complete the tasks following that instruction, nor have they considered what those may be.

5 ANALYSIS OF TASK 3: MEASURMENT OF MASS OF A SOLID

Introduction to the Task

Task 3 in these interviews was to measure and record the mass of about 0.5 g of a solid (sugar). Students should have, per the rubric, used a scoopula and either weigh paper or a weigh boat to place solid onto their intermediate on the balance. Any excess solid should have been removed to another utensil or directly into the waste, to avoid contamination. Upon taring the balance and reading the final mass of solid to record, interviewees were instructed to close the balance doors in order to avoid drafts from around the room causing variation in the balance reading.

Table 5-1: Rubric for Determining Mass Traditional rubric

Exemplary (5)	Cleans first uses weigh paper/boat + scoopula proper tare closes door waits for stable value records all digits + unit
Acceptable (4)	doesn't brush clean doesn't wait for stable value All other from exemplary
Neutral (3)	doesn't close door, or leaves off digit/unit, or both doesn't brush and doesn't wait improper tare
Poor (2)	two from neutral
Very Poor (1)	three or more from neutral

Although the initial rubric was as shown in Table 5-1, above, and included brushing existing solids from the balance before and after using the balance, this was a behavior that clearly was not enforced in these labs, since no participants completed this

step. Since not a single participant brushed the balance, students were not score penalized for this step (although it appears on the rubric) – since this is not discriminatory, it makes a frivolous entry on the rubric.

Discussion of Surveys and Interview Behaviors

In the pre-interview survey, students perceived that that determining the mass of a solid was overwhelmingly easy: 37 of the 38 students (97%) rated the task as easy or very easy (M = 4.63, SE = 0.11). These perceptions were largely unchanged in the post interview survey: 33 out of 35 students for whom data was available ranked this task as easy or very easy, (M = 4.51, SE = 0.12). This did not represent a significant difference in pre/post interview survey scores.

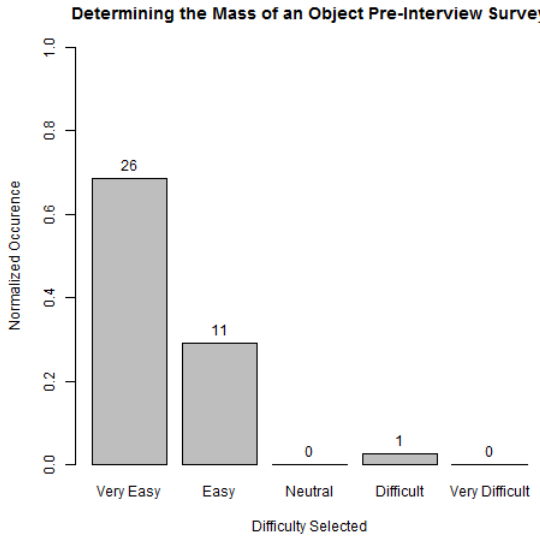

Figure 5-1: Pre-Interview Survey Responses

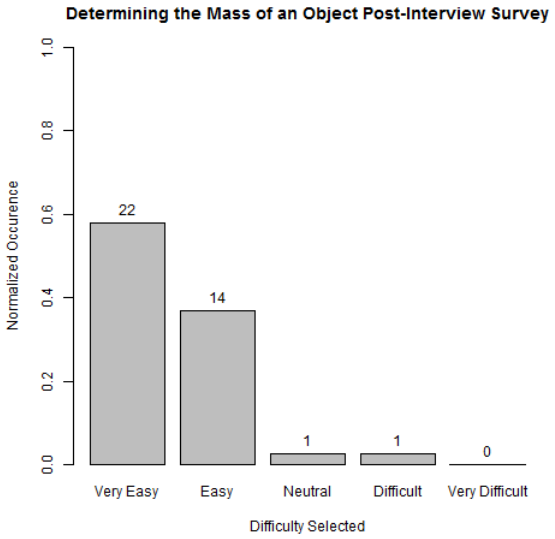

Figure 5-2: Post-Interview Survey Score

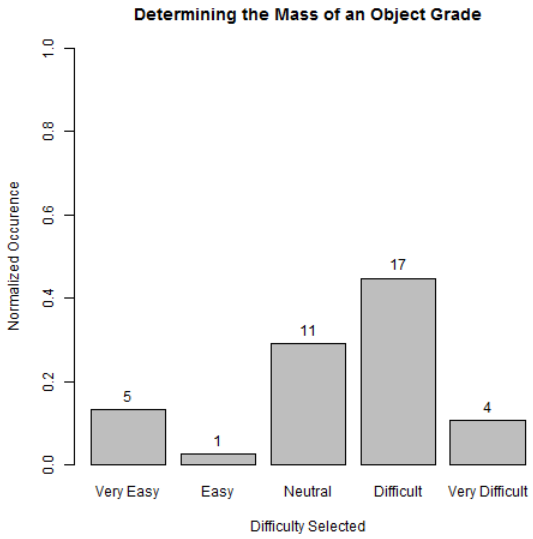

Figure 5-3: Interview Grade

Students were expected to perform this task handily overall – it was rated in surveys to be very easy, and many students use kitchen scales in their every day (outside of chemistry lab) lives. The researcher did expect some portion of the interview population to leave at least one door open while recording their mass, and for some students to not record their mass to all readable digits or to neglect to record the unit.

It happened that 18 out of 38 interviewees (47%) recorded the mass with open balance doors, which indicates that we as a teaching staff are not emphasizing enough what the doors are there for. Especially in a room with 15-20 other people walking around, it is important to close off a balance to drafts or other disturbances to the balance's reading, if students are expected to trust all digits recorded.

Another dire finding from this set of interviews was that of the 38 interviewees (all interviewees completed this task) 25 of them, or 65% overall, removed solid from the balance/paper/boat and replaced it into the original container. This observation is

shocking for two reasons: the first being that with an incidence of 65% in a relatively small sample size, the likelihood of this being a false positive observation (something that isn't happening in the population at large to any significant degree) is slim to none. Therefore, there is absolutely no question that a majority of the general chemistry students exhibit this behavior. The second implication of this observation is that students are not being appropriately warned of the dangers of contamination, or they do not realize that their actions are directly leading to contamination as it stands. In either case, this problem could lead to unusably contaminated reagents that are not appropriately apportioned, and could thus lead to incredible amounts of waste. Since many of these students are relatively familiar with kitchen tasks, it would be a worthwhile comparison to make them think of using the same spoon to put chopped garlic or brown sugar/honey (something pungent or something sticky) from a jar into a pan and then either back into the jar from the pan (most direct application) or then into another jar with the same spoon. In either case, the contamination is immediately recognizable thus helping students to understand the gravity of contamination woes (not so much individually but as a collective whole).

Less widespread problems included 9 students who did not record the mass at all, and 7 students who recorded the mass but not the unit. This means that 14 students out of 38 did not record the unit (grams) of their data. Particularly if we are training these students to become medical professionals, we need them to understand the importance of recording units – if they send prescriptions to pharmacies without units the results could be truly disastrous.

A noted finding without much weight was that of the 38 students interviewed, 21 selected a weigh boat to measure out their solid into, while 13 used weigh paper and 2 used a filter paper. While weigh paper and weigh boats are essentially interchangeable since both are inert and smooth to any substantial degree of granularity of solid, filter paper is explicitly rough and would thus catch a significant amount of solid thereby reducing the amount of solid transferred to the reaction by an unknown amount. This would not be a problem necessarily if students who used filter paper tared the balance (or recorded the mass of their filter paper) initially and then subtracted that mass from that of their filter paper after transferring their solid, but neither student who used filter paper in this interview completed that final step.

Also interesting to the researcher but of little consequence ultimately: 12 out of 38 students did not record a leading zero on their mass. For hand written measurements, the leading zero makes explicit that the number in question is less than 1 gram. even if the decimal point is less obvious,

Finally, 3 students (a little less than 8% of the population) added the solid on the counter, rather than inside the walls of the balance. Since the solid for this interview was table sugar, this struck the interview team as strange. There was no concern of corroding the balance itself, and spilling solid did not appear to be a major concern of any interviewees. A little further analysis of the lab manual students were using shed light on this behavior: In lab 3 of the 1211 lab manual, students are asked to add the solid iodine on the counter, rather than directly over the balance, because this would reduce the risk of spilling solid iodine on the balance and thus corroding them. The research team discussed possible solutions to this problem, since the consistent moving back and forth of the (not

all that sturdy) paper/boat increased the likelihood of losing solid on the counter. One possible solution is to get liners for the balances (they are like parafilm) to prevent having students add the solid iodine on the counter. Another solution would be to further emphasize to students in the instructions for the zinc iodide lab that this is a special scenario because of the caustic reagent, and that this is not the usual process for transferring quantitative amounts of solids. Ultimately, while this behavior was strange, it wasn't all that detrimental to interviewees' overall process so enough time and energy has already been devoted to solutions.

Time on Task

Time on task data for task 3 is displayed on page 42, in Figure 5-4.

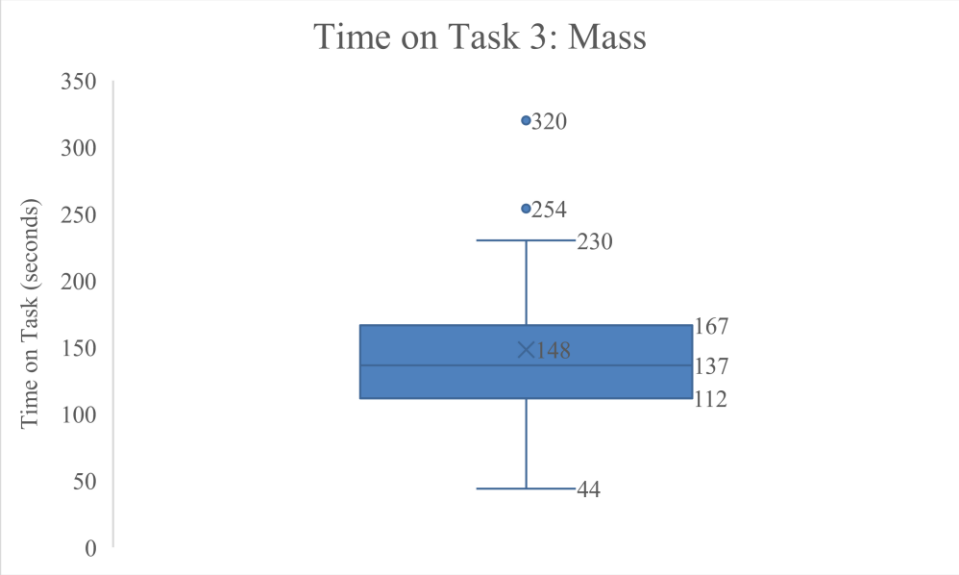

Figure 5-4: Time on Task (seconds) box and whisker plot for task 3

Average time on task was 148 seconds (about 2.5 minutes) while most interviews fell between 44 and 230 seconds (about 1 – 6 minutes). Half of all interviews took

between 112 – 167 seconds, or between 2 and 4 minutes. Given that many students over allocated solid to the balance before removing portions, this was within the realm of expectation for task completion.

Discussion of Transcript

The eleven most commonly used words in this task are noted in Figure 5-5, on page 44. These terms are, alphabetically: and, get, going, grams, just, little, mass, paper, put, solid and sugar. While filler words (stopwords) “and” and “just” appear in this list, overall this is a scientific/naming heavy list of top terms. Those scientific terms were: grams, mass, paper, solid and sugar. This indicates that students were being specific (i.e. solid or sugar over it or stuff) when describing their actions. The action words in the list (get, going, put) were also a heavy presence. The most used term was far and away going: with 95 occurrences, the word going was used more than twice as many times as any other term in this task. With 343 distinct terms in task 3, occurring 1444 total times, this means that the word going accounted for 6.6% of all words said during task 3.

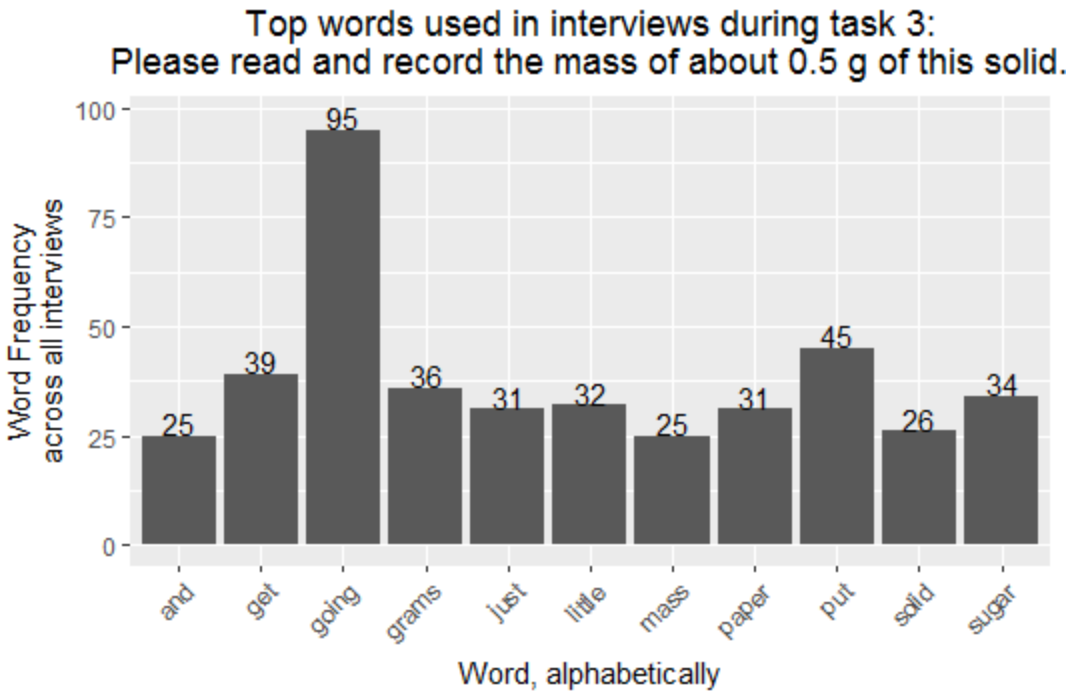

Figure 5-5: Histogram of 11 most frequently occurring terms in Task 3

To put the eleven most frequently occurring terms into perspective of the other frequently used terms in the third task, the 50 most commonly occurring terms are plotted below, in Figure 5-6. While those eleven terms certainly dominate the image, other emergent themes include the size of weigh/weighing/weight, boat, scale, singular gram, the word half, tare/zero, scoopula. These items occurrence means several things: first that stemming would have made a sharp difference in the appearance of these figures. Second, that students are certainly being specific in their descriptions of their actions. While less specific terms certainly are present in the cloud: thing, little, bit, those terms' sizes in the cloud is substantially smaller than that of the more specific terms.

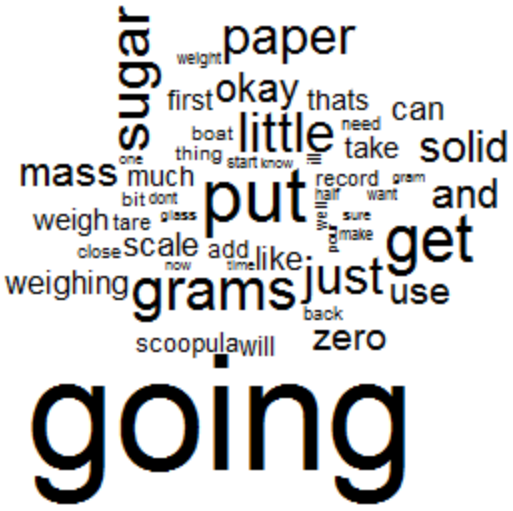

Figure 5-6: Wordcloud of 50 most commonly occurring terms in Task 3

Conclusions from Task 3

While users exceeded researchers' expectations for use of correct materials to measure solids on, there were still two participants who completed this task on filter paper. Most participants did use either weigh paper or a weigh boat to measure their solids. An alarming portion of the interviewees did not appear to be concerned with contamination via replacing excess solid into original containers, which is an indicator that further training is needed for these students regarding contamination awareness. Participants spent a perfectly reasonable amount of time on this task, nothing unexpected in that respect.

The mismatch between perception and performance on this task appears to almost entirely boil down to lack of contamination awareness throughout the class, and

neglecting to close balance doors during taring and recording masses. To support this theory, one could re-score those interviews with a rubric which did not deduct points for contamination behaviors or neglecting to close balance doors, and test for significant difference in perceived difficulty versus scores. If there was not then a significant difference in perception/performance, this would support the hypothesis that the mismatch in perception/difficulty for these interviews boiled down to those two problems.

6 ANALYSIS OF TASK 4: TRANSFER 10.00 ML OF SOLUTION

Introduction to the task

The fourth task in the observations portion of the interviews was transferring 10.00 mL using either a graduated (Mohr) pipet or volumetric pipet from one beaker to another. The prompt read: "Transfer 10.00 mL of this solution to beaker D from Beaker C." The method of transfer was left vague intentionally: the research team was interested in whether students suspected preference for using graduated cylinders over pipets even when pipets are a more logically sound choice, would hold in interviews (rather than just anecdotal evidence). An unexpected problem with students' performance of this task was reading and comprehension of the task itself. While most students read the task as it was written, many 15 struggled over the to/from format of the sentence in the prompt – even after reading it correctly, sometimes multiple times[27-29]. Perhaps if this research is replicated with new interviews, researchers should switch from using to/from structure to from/to structure, since there was such a visible problem with the sentence structure.

Table 6-1: Rubric for Use of Volumetric Pipet

Exemplary (5)	Is using volumetric pipet Checks fit of pipet in holder Tests with small amount of liquid Slows at wide part
Acceptable (4)	Uses volumetric pipet but doesn't check fit or test small amount
Neutral (3)	Is using graduated but uses correctly Doesn't slow at wide or overshoots line but gets back down to line
Poor (2)	Uses graduated at 4 level or Uses volumetric and doesn't check fit/test small amount and doesn't slow
Very Poor (1)	Uses graduated but still wrong Overshoots/undershoots

Table 6-2: Rubric for Use of Mohr (Graduated) Pipet

Exemplary (5)	Checks fit of pipet in holder Tests with small amount of liquid Slows as approaches 10, 0 mL lines Stops at 10/0 mL line at bottom of pipet
Acceptable (4)	Stops at 0 mL line and one of: Overshoots 10 mL line at top of pipet but rolls back down Doesn't test Doesn't check fit
Neutral (3)	Stops at 0 mL line and two of: Overshoots 10 mL line at top of pipet Doesn't test Doesn't check fit
Poor (2)	Goes past 10/0 line but expresses that shouldn't have
Very Poor (1)	Goes past 10/0 line with no address

Table 6-3: Rubric for Task 4, Liquid Transfer

Exemplary (5)	Uses Mohr or volumetric pipet unprompted at 5 level
Acceptable (4)	Uses Mohr or volumetric pipet unprompted at 4 level
Neutral (3)	Use of pipet unprompted at 3 level
Poor (2)	Use of pipet unprompted at 2 level
Very Poor (1)	Use of pipet unprompted at 1 level
Requiring a prompt to complete this task via pipet (first completing via graduated cylinder etc.) is a 1 point deduction from performance, and is dependent on use of other glassware	

Discussion of Survey and Interview Data

Pre- and post- interview survey distributions of responses to difficulty of use of a transfer pipet are found in Figure 6-1, below. In the post interview, only 16 of the participants selected very easy for use of transfer pipet, but 34 students selected easy or very easy, (M = 4.45, SE = 0.09). This did not represent a significant change from those interviewees in the pre-interview survey. The survey item using a volumetric pipet (M = 4.09 SE = 0.13) decreased slightly, but this did not represent a significant decrease in difficulty, t (35) = 0.723, p = 0.475.

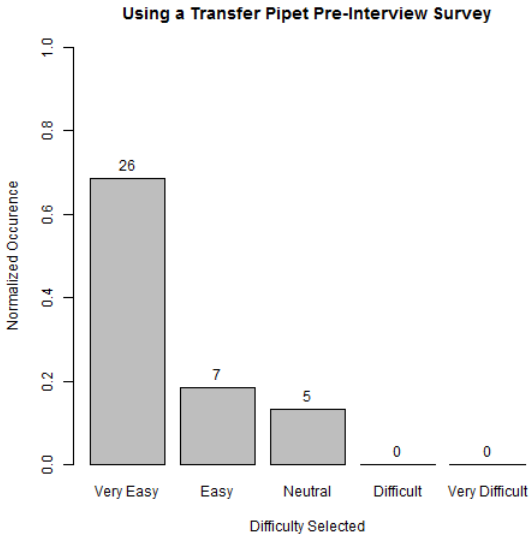

Figure 6-1: Pre-Interview Survey, Using a Transfer Pipet

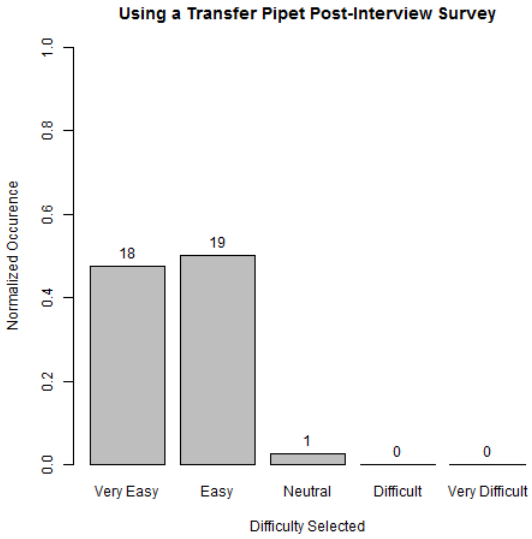

Figure 6-2: Post-Interview Survey, Using a Transfer Pipet

Survey items relevant to this task were: Use of a volumetric pipet, use of a transfer pipet, and selection of proper glassware. Glassware has already been analyzed in the Task 1 chapter of this document. In the pre-interview survey, 33 students out of 38 perceived use of a transfer pipet as easy or very easy, (M = 4.55, SE = 0.12), but 26 of those 33 selected very easy. By comparison, 30 students thought that use of a volumetric pipet was easy or very easy, (M = 4.21, SE = 0.21), 15 of those 30 selected very easy. Volumetric pipet pre- and post-interview responses, detailed in Figure 6-3, Figure 6-4, on page 50, were among the very few sets of responses that were very different from pre- to post-interview.

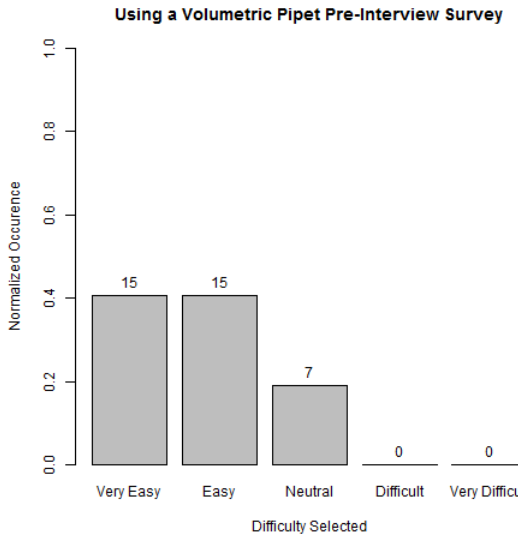

Figure 6-3: Pre-Interview Survey, Using a Volumetric Pipet

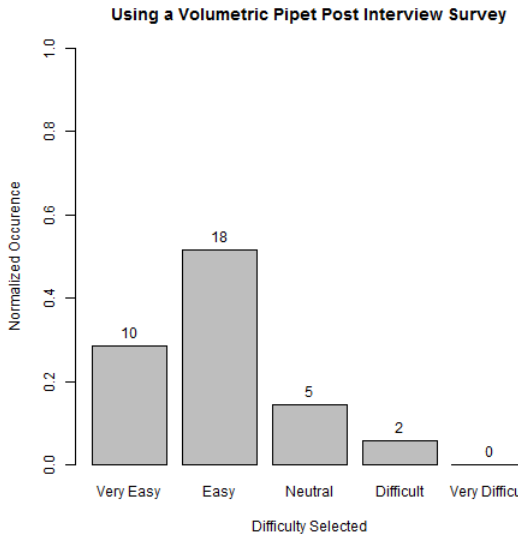

Figure 6-4: Post-Interview Survey, Using a Volumetric Pipet

Of the users who initially rated using a volumetric pipet as very easy (5), 8 rated the task still as very easy in the post-interview survey. The remaining participants who initially rated using a volumetric pipet as very easy changed their response in the post survey, but all changed their responses to easy (4). One participant who initially responded that this task was easy (4) changed their response in the post survey favorably, to very easy, two changed unfavorably, to neutral, and the remainder stuck with easy. Of the initial seven who responded that this was a neutral task, neither easy nor difficult, three changed favorably, 2 to easy, one to very easy, and one unfavorably to difficult. This points to users not changing perceived difficulty from pre-interview to post-interview much overall, and only moving one level of difficulty when they do overall in either direction. These analyses of change in perceived difficulty necessarily excluded participants missing either pre- or post- interview data (ie skipped back page of pre-interview or selected not applicable).

Although it was discussed in a prior section of this paper, selecting glassware pre- and post-interview survey responses are included below, in Figure 6-5 and Figure 6-6.

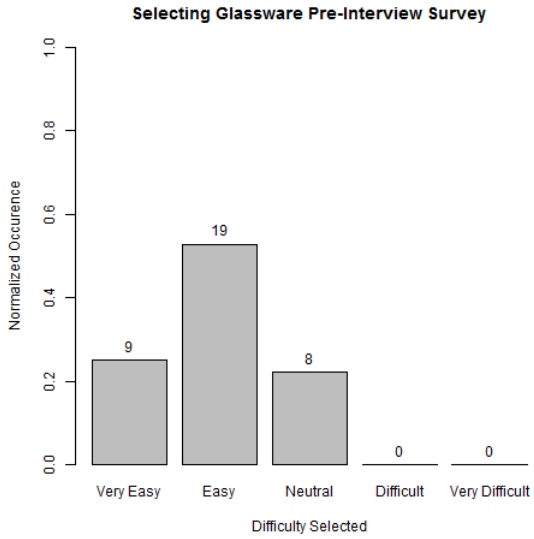

Figure 6-5: Pre-Interview Survey, Selecting Glassware

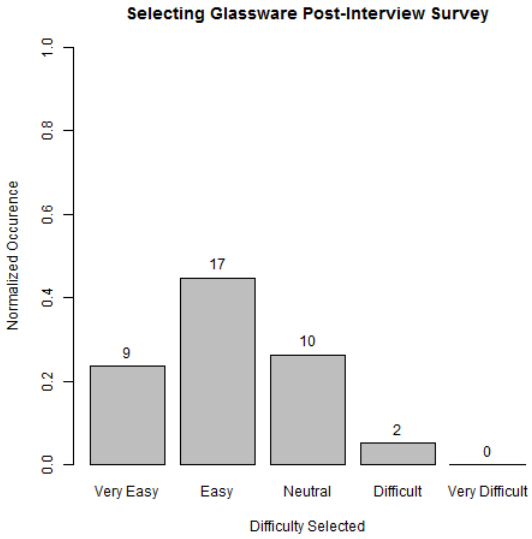

Figure 6-6: Post-Interview Survey, Selecting Glassware

Finally, scores from the rubric on using a volumetric pipet (for those participants who did use a volumetric pipet during the interview process) is detailed below, in Figure 6-7. This was a near even distribution with a slight skew toward easy, with 17 participants scoring a 4 or 5 on use of a volumetric pipet, out of a total of 34 who used one.

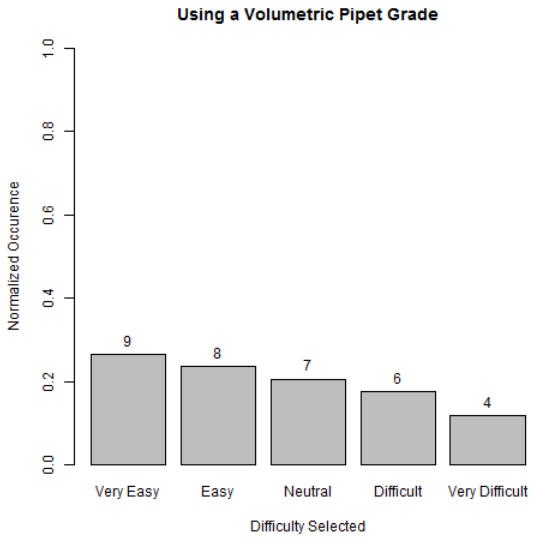

Figure 6-7: Using a volumetric pipet score

Use of a volumetric pipet grade (M = 3.35, SD = 1.37) was lower significantly than either pre- or post- interview surveys, however this difference could just as easily be attributed to problems with the rubric with students' learning patterns, as with discrepancies between students' perceived and actual skill levels with this task. One way to discriminate between these possibilities would be to perform an inter-rater reliability study on the rubric for use of a volumetric pipet grade. This would either confirm the validity of participants' grades on task performance, leaving the conclusion that students' perception of using a volumentric pipet is that it is significantly easier than their performance of the same task; or would confirm that the rubric requires reworking to appropriately assess students' performance of this task.

Discussion of common behaviors in interviews

Of students who read this task (all 38) 14 (37%) initially used the small (10.0 mL) graduated cylinder instead – which, while technically correct, is not the best selection of glassware to complete this task. A small, non-negligible amount of liquid remains in the bottom of the small graduated cylinder, and that amount is not specified from cylinder to cylinder. To-Deliver pipets, meanwhile, are calibrated to leave a small amount but still to transfer their specific amount, when used properly. That when used properly disclaimer is an important step. The lab manual doesn't include instructions (in either 1211 or 1212) on use of the green winding pipet pumps which are used throughout the course. The instructions only describe use of bulb pipets without stopping capabilities, that require removal and don't allow for fine adjustments at all. After filling past the line, the instructions call for gross adjustments by removing the bulb, and using your thumb to release small amounts of liquid to get the meniscus to the line.

Since the chemistry laboratories have ample green pumps for use, , and since they are used by both classes throughout the series, there needs to be a set of clear, illustrated instructions on use of green winding pipet pumps in the student manual. Because those instructions do not exist, there were 3 students who removed the winding pump in the observations. The researcher did not ask why at the time of the interviews, but further examination of the lab manuals found the discrepancy and lack of instructions on the winding pump. Also relevant to better training on use of pipets: 4 students visibly from video playback touched the tip of the pipet to the bottom of the beaker. 10 visibly checked the fit of the pipet within the holder, and 2 tested the fit of the pipet within the holder with a small amount of solution.

Overall, 13 students used a pipet correctly (either Mohr or volumetric). 12 students used a Mohr pipet (this is both initially and after being prompted to use a pipet) several calling the Mohr pipet a volumetric pipet. Of the 12 who used a graduated pipet, 9 filled to the 0.0 line, and then dispensed past the 10.0 line, rather than stopping at the line. Of the students who used a Mohr pipet, they still could have gotten full credit (5 points) on the transfer task had they stopped on the 10.0 mL line. Finally, although all 38 students were presented with this task and prompted, 4 students never used either a Mohr or a volumetric pipet in this interview.

The lab manual does not sufficiently explain that for small volumes, especially volumes for which students have volumetric/graduated pipets available, pipets should be used for straight transfer of solutions. This can in part be attributed to purchase of Dispensette Pumps for general chemistry lab, which have reduced contamination and waste within general chemistry labs. This can also be partly attributed to poor instruction on winding pumps within the manual. Overwhelmingly, students were most comfortable using graduated cylinders for transfer of small amounts of solution, which is problematic since trace amounts of their solutions are left in the cylinder, without calibration for or a way to calculate how much that is.

Discussion of Transcript Information from Interviews

The ten most commonly used phrases during this task are detailed below, in Figure 6-8. These terms were, alphabetically: beaker, cylinder, going, just, like, okay, pipet, solution, transfer, and use. The most frequently used term was going, with 57 total occurrences, followed by beaker with 47. Since the intention of this task was to get

students to use a volumetric pipet, it was interesting to note that cylinder (as in graduated cylinder) made this top ten most frequently used terms list, with 17 occurrences. It was outpaced by 24 uses of pipet, which is heartening.

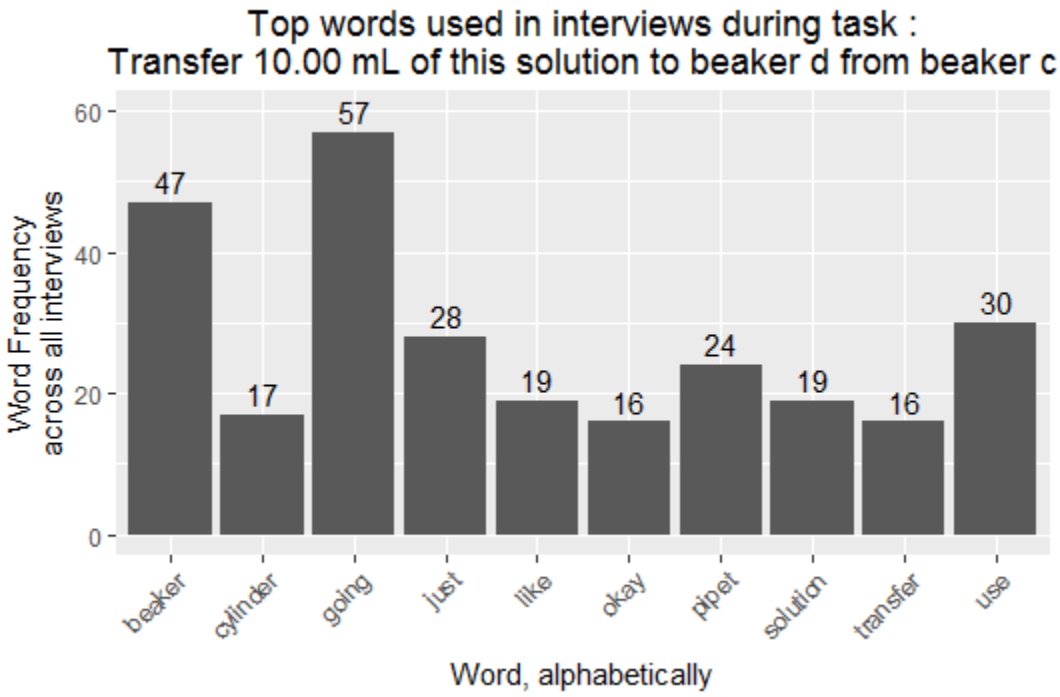

Figure 6-8: Histogram of 10 most commonly used terms in Task 4

From 020201:

> "So um, I have a solution of some sort in beaker c and I have nothing in beaker d. So what I would do is I will get the uh, the 10 mL graduated cylinder and I would basically pour a little bit of the solution into this, and uh once I get close I kind of eyeball it so I can get the uh, the 10 mL range which I did, but that can easily be done by pouring a little bit back until you have approximately 10 mL of a solution. Which is really really close, if you have to you can use a little pipet to like make it exactly 10, but that will depend on, once you have that you just pour it into beaker d, and that's 10 mL of the solution in beaker d."

This excerpt from 020201's interview exemplifies the casual language many students used throughout the interviews, but also is a good example of an interview in which the user used a graduated cylinder, without use of a transfer pipet, but with the mention of using

one to get to an exact measurement. This does show that this user was cognizant of how to obtain a more accurate measurement, and that the participant was considering significant figures available to them and simply decided not to perform that action.

In Figure 6-9, below, is a wordcloud of the 50 most frequently occurring terms from this task. Something this wordcloud does not share with other task specific wordclouds is that this one has a few extremely dominant terms, beaker and going, and then a plethora of other terms (many more than 10) all near the same size, indicating similar numbers of occurrence, before the smaller set of tiny terms which occurred least frequently amongst the top 50 terms.

Figure 6-9: Wordcloud of 50 most commonly occurring terms in Task 4

Those medium-sized words which didn't make the top 10 list include meniscus, graduated, volumetric, pour, line, exactly. This indicates that both the researcher was correct in assuming cylinder was referring to graduated cylinder, and that users were

again being specific in their descriptions of their actions. Participants were more likely to use names of their equipment than vaguer words like it or thing.

Conclusions from Task 4

Although the intention of this task's inclusion in the study was to force users to use a volumetric or Mohr pipet, and to test use of these pipets, and whether students understood the difference in these two pipets, these goals were not ultimately accomplished by the task as written. In the first couple interviews, the task prompt read "…with a volumetric pipet" since this was an item that the researcher wanted to test with this interview. However, with the changes to the prompts that occurred after a few interviews had been performed, the research team discussed and decided that it would be interesting to see if, when the specific use of a pipet was not requested, students understood that this amount of solution would best be transferred using a pipet. As it turns out, they (the students) did not all understand this distinction. The changes to the prompt were also followed by follow-up questions at the end of the interview, asking users to identify and select a volumetric pipet, and use that piece of glassware to complete the task again, so that the volumetric/graduated pipet distinction could still be observed. There were still 4 participants who were not observed using a volumetric or graduated pipet, despite prompts, due to either time constraints or inability to complete the task with their selected glassware. Of the 34 students who did complete a transfer with a pipet, 12 of those used a graduated pipet, rather than a volumetric pipet. Since completion of the task with a graduated pipet (when used correctly) still had the potential to get full marks on the task, users who selected a graduated pipet unprompted to

complete this task were never prompted to select a volumetric pipet. For this reason, we cannot draw the conclusion that none of the 12 students who did complete the task with a graduated pipet could not correctly identify a volumetric pipet. An introduction to glassware and appropriate uses in their lab manual, perhaps before the first lab, could eliminate some of this confusion about when to use which glassware, and what each piece of glassware is called.

7 TASK 5: TRANSFERRING LIQUID USING A BURET

The fifth task in this interview/observation session was to transfer 10 mL of solution with a buret: “Use the buret to transfer 10 mL of solution to a beaker.” In an effort to not prompt interviewees to record with the appropriate number of digits, or even to prompt them to record values at all, “10” was left as a whole number with no decimal places. For perfect performance of the task, students were expected to record both the initial and final volume readings in the buret (actual, not anticipated), to read the buret correctly, and to slow down to a dropwise pace as they neared the anticipated final volume. The buret was not left at a controlled height throughout the interviews, and thus was too tall for some students and appropriate height for others. Some students physically lifted the buret within the holder to place the receptacle beaker underneath the spigot, but did not then replace the buret to a lower height (so that the distance was appropriate between the tip of the buret and the level of the liquid). To not be dinged for the space too much error, students needed to have about 3 cm or less between the stopcock and the top edge of the receptacle beaker.

Table 7-1: Rubric for Reading a Buret

Exemplary (5)	Glass height adjusted for student Numbers toward student Reads from bottom of meniscus Records volume to +/- 0.03 mL
Acceptable (4)	Missing one from exemplary
Neutral (3)	Volume to +/- 0.07 mL doesn't record unit
Poor (2)	Missing 2 from exemplary Both from neutral
Very Poor (1)	Volume more than .07 mL off and does not record unit. Does not read from eye level. Glass not height adjusted.

Table 7-2: Rubric for use of a Buret

Exemplary (5)	Reads properly. Turns stopcock slowly, only allows a drip not a full stream. Slows speed near final volume dispensed. Ensures stopcock closed before filling.
Acceptable (4)	Can't read properly because of height Full stream at first but slows before anticipated point
Neutral (3)	Two of: stopcock open to fill but otherwise 5 use. allows stream but still stops at anticipated point. 3 use of reading, but otherwise 5 use of buret. over/undershoots anticipated volume dispensed
Poor (2)	Neutral use AND Height adjustment OR Reading at 3 level AND acceptable
Very Poor (1)	Combination of 2 or more from neutral. OR reading at 1 level.

Discussion of Survey and Interview Data

In the pre-interview survey, interviewees thought that reading a buret and using a buret were both easy tasks, with reading a buret (M = 3.91, SE = 0.14), while using a buret was (M = 3.91, SE = 0.13). These survey responses can be seen below, in Figure 7-1 and Figure *7-3.*

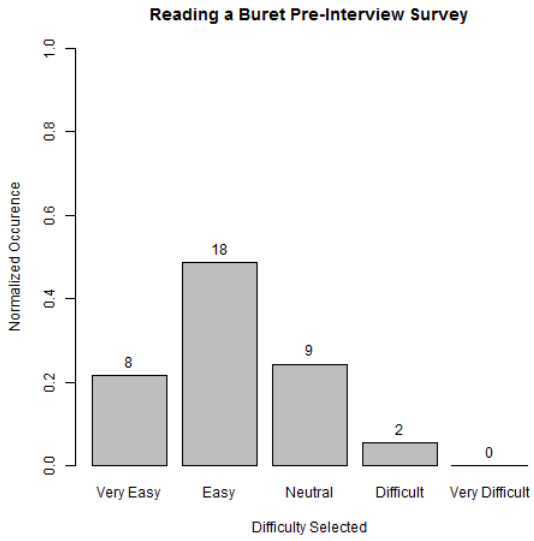

Figure 7-1: Reading a Buret Pre-Interview Survey

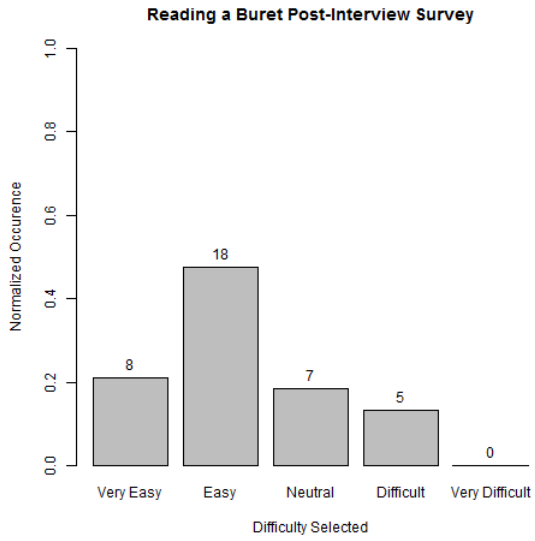

Figure 7-2: Reading a Buret Post-Interview Survey

Of the initial 8 participants who selected very easy for reading a buret in the pre-survey, four adjusted their responses in the post-survey unfavorably: 3 to easy and one to difficult. The remaining four stuck with their responses of very easy. Of the 18 participants who selected easy in the pre-survey, 4 adjusted their responses favorably and 2 unfavorably. The two unfavorable adjustments were both to difficult. The remainder of initial easy selectors stood by their initial responses. Of the nine participants who initially responded that reading a buret was neutral, neither easy nor difficult, 3 adjusted favorably and 1 adjusted unfavorably. All those who adjusted favorably selected easy in the post-interview survey, while the negative adjustment was too difficult. Finally, the 2

participants who initially selected difficult were split in their post responses, one adjusted positively to neutral, the other remained with difficult.

During the post interview survey, students thought that reading a buret was easy (M = 3.76, SE = 0.16) and using a buret was also easy (M = 3.92, SE = 0.12). Neither of these represents a significant difference from their respective pre-interview scores, t (36) = 0.726, p = .473; t (37) = -0.361, p = .720. This information can be found above and below, in Figure 7-2 and Figure 7-4.

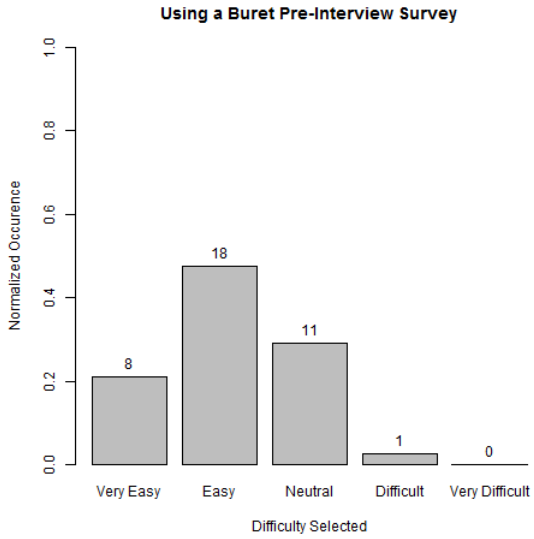

Figure 7-3: Using a Buret Pre-Interview Survey

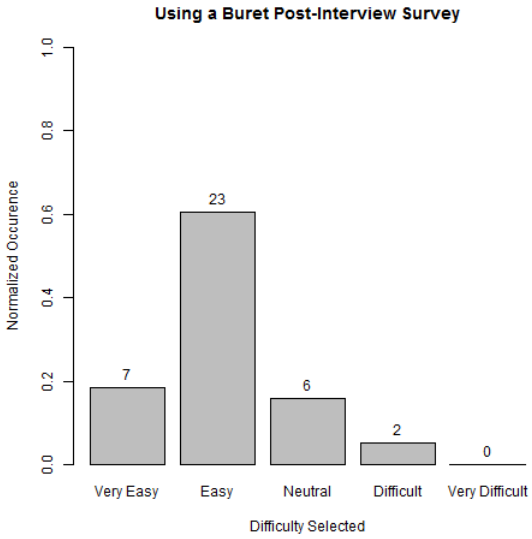

Figure 7-4: Using a Buret Post-Interview Survey

Of the initial 8 participants who selected very easy for use of a buret in the pre-survey, 3 responded the same in post-survey. The remaining 5 adjusted unfavorably from very easy to easy. Of the 18 who initially selected easy, 4 adjusted favorably to very easy, and four adjusted unfavorably, 2 to neutral and 2 to difficult. The remainder did not

change, and selected easy for both the pre- and post-interview survey difficulty. Of the 11 who initially selected neutral, seven adjusted positively to easy, the remainder did not change their responses. The sole participant who initially responded that using a buret was difficult changed their mind, to say that it was easy in the post interview.

While neither of these survey items represented a significant difference in value from pre-interview survey to post-interview survey, it was interesting to note the shift in shape of responses from pre-interview to post-interview. In the pre-interview survey, both items' responses were roughly normal, slightly easy-skewed (positively skewed, on these histograms) while on the post-interview survey, items became much less normally distributed with much more leptokurtic appearances (they were sharper, with more responses in the easy category and less/same in both very easy and neutral). The descriptive statistics essentially shook out the same way, but many students migrated to a perception in which the score of each item was easy from neutral, very easy.

Discussion of Common Behaviors in Interviews

Of the 38 interviewed students, 6 (16%) left too much space between the buret tip and the receptacle beaker. Similarly, 8 (21%) did not appropriately height adjust the buret for themselves to read the buret.

16 interviewees, or 42%, recorded at least one of the buret readings unprompted. Five students, or 12%, were prompted and then recorded the final volume in the buret. 8 did not record the initial volume (and were then prompted to record the final volume). 16 did not record the final volume: meaning, 16 did not record the actual final volume. Because the buret is recordable to two decimal places, students could not simply record

the anticipated final volume and not write the actual final volume. One student did attempt to remedy this by recording the anticipated final volume +/- 0.05 mL – this was reasonable attempt to remedy, though recording the actual final volume would have been better scientific practice. Eight of the students interviewed did not record the unit – again 21% of students interviewed don't seem to understand the absolute importance of reporting units with every single number they report. Less important but more widespread: 16/38 did not record the appropriate number of significant figures on at least one of the values reported. Although this concept is gone over several times within their lab manual, even specifically for buret readings, they aren't getting it.

By not recording actual final volumes, interviewees are falsifying data, though the researcher doesn't think that students understand that implication of their actions. This is indicative of a larger problem wherein students don't fully understand what constitutes honor code violations and cheating. This could even have broader implications of an entire generation of students who don't understand the ethical implications of doing and reporting science, and the importance of clarity and conciseness in their reporting. these Some students manifest a fundamental disconnect between their reported values, gathering of their own data, and proper citation of data acquired by others. This fundamental disconnect could possibly be addressed by adding a snippet into the first week lab of 1211, in which students gain familiarity with significant figures and data collection. Adding a unit in which they are quizzed on data attribution and correct reporting would be beneficial.

Time on Task Analysis

Time on task analysis was carried out for task 5, and is demonstrated visually in the box and whisker plot, below, in Figure 7-5.

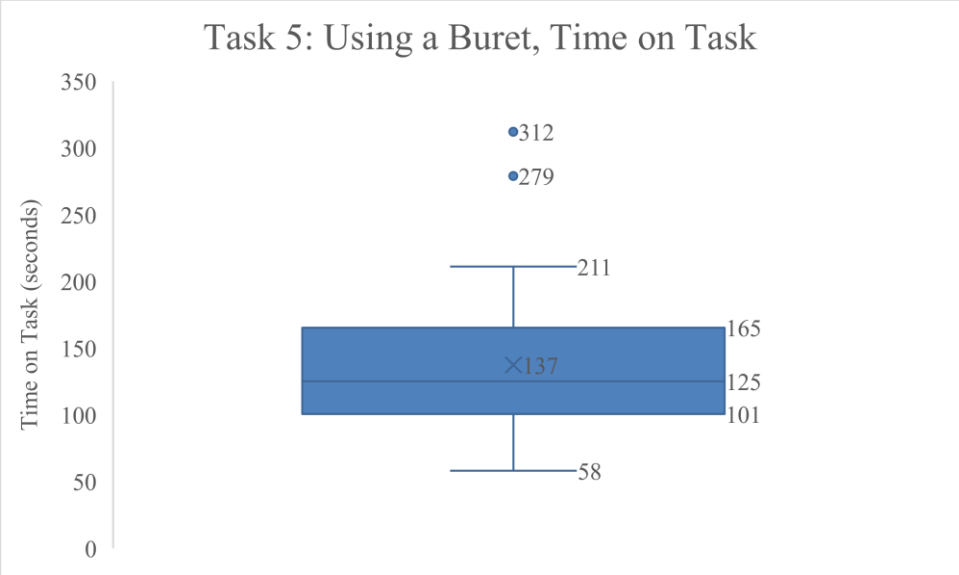

Figure 7-5: Time on Task (seconds) 5, box and whisker plot

While the majority of students completed this task between 58 – 211 seconds (1 – 4 minutes), there were a few who took longer – the longest was 312 seconds (a little more than 5 minutes). Average time on this task was 137 seconds, with half of interview participants taking between 100 – 165 seconds (1.5 – 3 minutes).

Discussion of Transcript Information from Interviews

Transcripts from this task were analyzed as the other tasks throughout this experiment, with a histogram of the eleven most frequently occurring terms in the task (Figure 7-6, below) and a word cloud of the 50 most frequently occurring terms. Those eleven most frequently occurring terms were, alphabetically: beaker, buret, get, going, just, like, little, okay, read, right, and that's. Although other tasks' most frequently used

terms indicated that very specific, scientific terminology was dominating the conversation about that task, this was not as visible with this task. There were scientific terms in the list (beaker, buret) but mostly this was a collection of action and filter words.

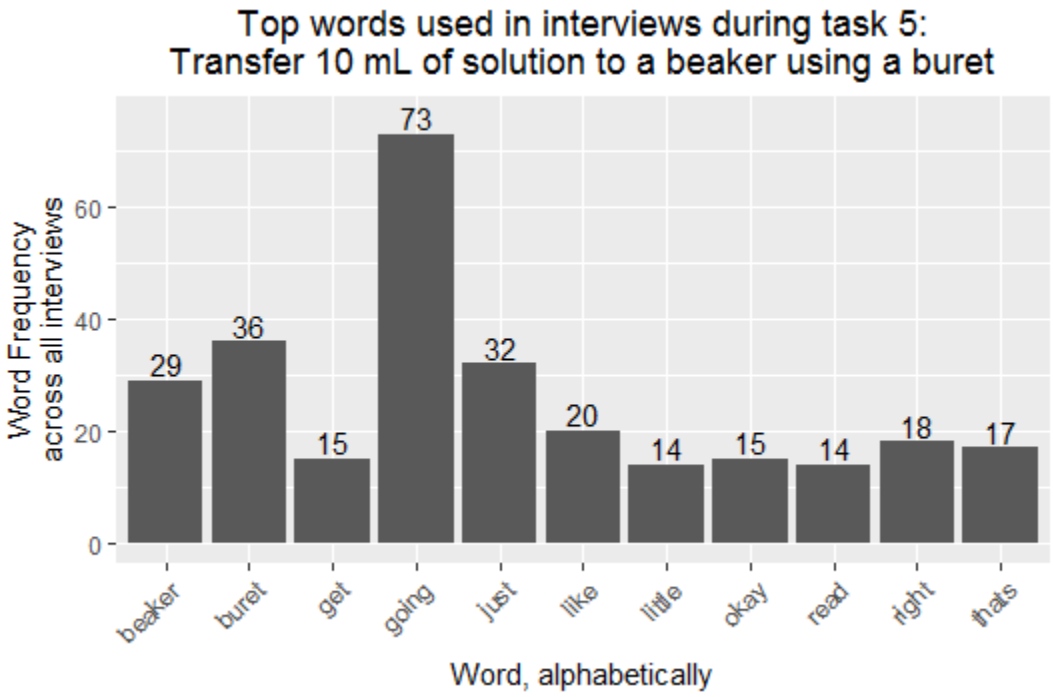

Figure 7-6: Histogram of the 11 most commonly used terms in Task 5

The word going occurred 73 times throughout task 5, which out of 971 total uses of 297 distinct terms is very impressive: that's saying that the word going represented 7.5 % of all words used in this task. It's important to remember that these are occurrences of terms **not inclusive** of reading the prompts themselves, those were stripped from transcripts before these analyses were completed. Combining the word count with time on task, we get an average word density of about 7 words per second.

A broader look at the language being used in this task can be achieved by examining the wordcloud for this task, Figure 7-7. The words solution, transfer, reading, volume, initial, and final were all clearly legible on the wordcloud as well. This indicates

that this task was not left completely out of students' tendency to speak specifically and scientifically about their actions and experiences during the interview, they merely used action and filler words more frequently.

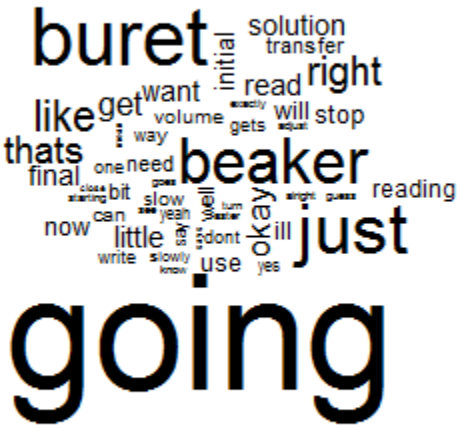

Figure 7-7: Wordcloud of 50 most commonly used terms in Task 5

There was a dearth of filler words in this cloud as a whole compared to other task-specific clouds, however. The ones which are exceedingly obvious on first glance at the cloud are: right, want, like, okay, just, little, bit, gets. Interesting to the researcher is that (although small) the word write appears, since many participants neglected to write initial, final, or both buret readings during task performance.

Conclusions from Task 5

The heaviest conclusion to be drawn from this task is that students do not fully understand the consequences of recording data as it is observed, rather than after the fact,

or not at all. It is of utmost importance that this point be emphasized to the student population as they complete the general and organic chemistry series here at UGA. An overwhelming majority of these students report pre-professional career aspirations and majors (pre-medicine, pre-veterinary medicine, etc.). The consequences of producing a cadre of doctors, veterinarians, dentists and lawyers unaware of the importance of recording observed, relevant data during their day-to day operations are staggering There is an enormous amount of training ahead of these preprofessionals, but by not making sure that these points are engrained in their educational foundation we are doing them a gigantic disservice.

Another conclusion with considerably less gravitas to be drawn from this task is that students may use less scientific/naming language, and more filler and action language, as well as less speech overall, when they are less confident in their prospects while completing a task. A way to test this in future studies would be to select students who have been only minimally exposed to techniques in a task or experiment, and ask them to complete that task in an active interview, as well as a very rudimentary task, also in active interview. With extensive transcription (as in these interviews) and comparison of the tasks' language, time on task, one could then draw conclusions about comfort with task complexity and amount/type of communication during interview.

8 ANALYSIS OF TASK 6: HEATING A SOLUTION

Introduction to the Task

The sixth task in the interview was heating a solution (water) to approximately 30°C. The task read: "Heat this solution to about 30°C". To get all 5 points on this task, students would have read and recorded the initial and final temperatures of the solution, selected a hot plate initially (not a stir plate) and plugged it in before beginning, and kept the thermometer off of the glass during readings, and set the hot plate to at least 50°C. While one could certainly set the hot plate to any temperature over 30°C to cause their solution to heat to 30°C, anything on the low side would cause the heating process to take a long time. Even the smallest alcohol (methanol) boils off before 100°C, but over 50°C. If students had been concerned about boiling off their solution before it heated to 30°C (I don't think they were) I would have told them that their solution was water, and would thus boil at 100 °C. I did ask a few of the 30/35°C segment why they set the temperature of the hot plate to that temperature, (and the higher temperatures too). The low (30/35) temperature participants did say variations on it said to heat it to 30/35, or I don't want to boil it off, but didn't make the connection that they wouldn't have boiled it off, or that increasing the temperature on the hot plate would only heat their solution faster not increase the solution temperature to what the hot plate read instantly. The participants who stuck to the low temperatures don't seem to have made the connection between hot plate temperature and speed of heating, or didn't appreciate the heat of vaporization – and

seemed equally distressed and confused that their solutions were taking so long to heat when set to such low temperatures.

Anecdotal expectations lead researcher to expect students to leave their plates unplugged and or select a stir plate rather than a hot plate The research team noted that 3 students (out of 38) selected the stir plate over the hot plate. The team also found that 4 students attempted to heat their solutions without plugging in their plate.

Table 8-1: Rubric for Use of a Thermometer

Exemplary (5)	Thermometer lifted off bottom of glass but fully within solution Read from eyelevel and normal to gaze Held in solution long enough to read temperature
Acceptable (4)	Not at eye level or Not normal to gaze
Neutral (3)	Thermometer in liquid but touching bottom glass or Not in substance long enough to tell temperature
Poor (2)	One each from 3 & 4
Very Poor (1)	Thermometer touches glass Thermometer not held in solution long enough to equilibrate Not at eye level Not normal to gaze

Table 8-2: Rubric for use of a Hot Plate

Exemplary (5)	Selects proper equipment. Plugs in before turning on. Does not turn on empty. After reaching temperature switches plate off. Cools/unplugs before putting away
Acceptable (4)	One of: Turns on empty Doesn't turn off plate after reaching temperature Doesn't cool off/unplug before putting away
Neutral (3)	Starts with stir plate but completes rest from 5 OR Doesn't plug in before heating but then completes after realizing mistake
Poor (2)	Both of 3 and one of 4
Very Poor (1)	2 or more of items from 3 or 4.

Table 8-3: Rubric for Task 6, Heating a Solution

Exemplary (5)	5 use of thermometer and hot plate
Acceptable (4)	4 use of one
Neutral (3)	4 use of both or 3 use of one
Poor (2)	3 use of both or 2 use of one
Very Poor (1)	2 or worse use of both

Discussion of Interview and Survey Statistics

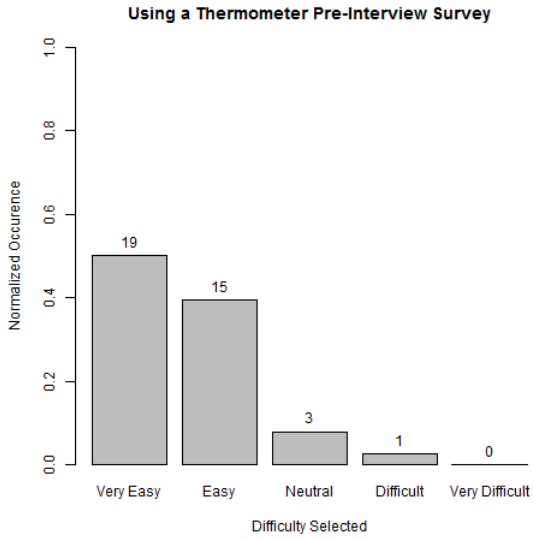

Figure 8-1: Using a Thermometer Pre-Interview Survey

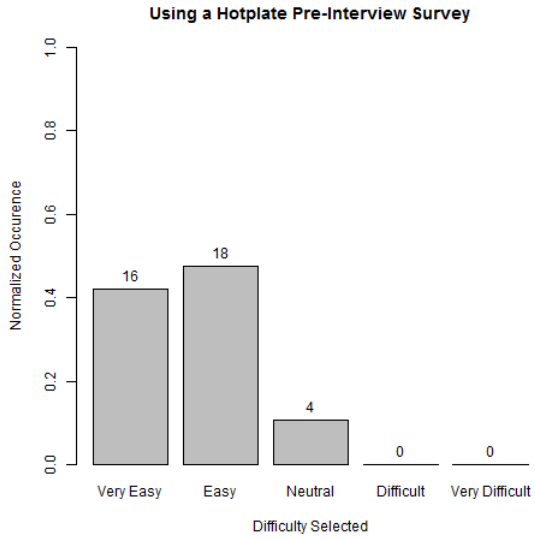

Figure 8-2: Using a Hot Plate Pre-Interview Survey

In the pre-interview survey, 34 students perceived using a thermometer as easy or very easy, (M = 4.34, SE = 0.13), while 34 students perceived using a hotplate as easy or very easy (M = 4.32, SE = 0.11).

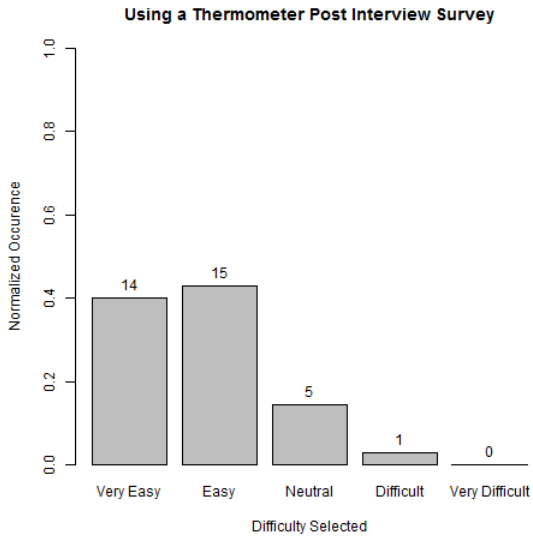

Figure 8-3: Using a Thermometer Post-Interview Survey

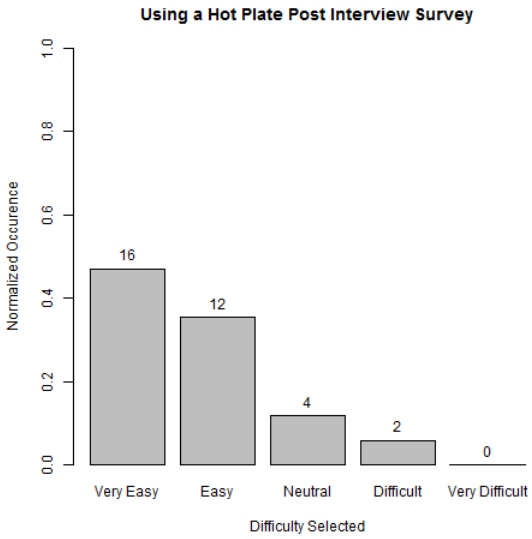

Figure 8-4: Using a Hot Plate Post-Interview Survey

In the post interview, however, there was the slightest shift to more difficult opinions: 29 participants selected easy or very easy for using a thermometer, (M = 4.20, SE = 0.14), while 28 participants selected easy or very easy for using a hot plate (M = 4.24, SE = 0.15) respectively, however these still fell within the realm of very easy tasks, with non-significant differences in their overall scores from pre- to post- interview scores, t (34) = 0.927, p = .361; t (33) = 0.414; p = .585.

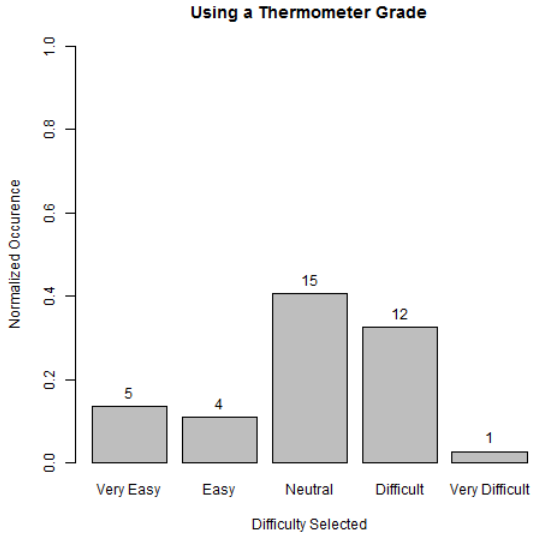

Figure 8-5: Using a Thermometer Interview Performance Score

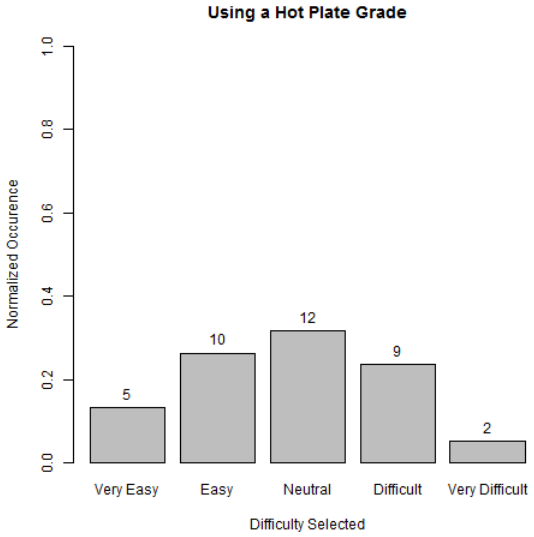

Figure 8-6: Using a Hot Plate Interview Performance Score

Grades for using a thermometer and using a hot plate were substantially less than scores on both pre- and post-interview scores, (M = 3.00, SD = 1.05), (M = 3.18, SD = 1.11).

Discussion of Common Interview Behavior

Of the 38 students interviewed, 24 used a thermometer, and of those 10 recorded the initial temperature unprompted, while 6 recorded the final temperature. Of these, 4 people recorded both initial and final temperature. 19 of those students (55.26%) who used a thermometer allowed their thermometers to rest on the glass – some even after explaining that allowing the thermometer to rest on the glass would measure the temperature of the glass, rather than that of the solution. One student decided that they needed a clamp of some kind to hold their thermometer, rather than holding it the whole time, letting it rest, or removing and replacing the thermometer each time they checked

the temperature, but said that they could not find the appropriate clamp available. They were offered both utility clamps and buret clamps, but said both would crush the thermometer.

Five students out of the 38 interviewed turned on the hot plate empty – which is explicitly warned against in the 1211 lab manual. Finally, 11 of the 38 students (28.95%) used some variety of personal protective equipment when dealing with the hot plate task. One student who didn't use personal protective equipment mentioned that they had grown up in Europe, and thus had a substantial background to know that 35^{o}C is not hot – specifically they mentioned that it is less than body temperature so they were not worried about burning themselves or the counter.

Time on Task Discussion

Time on task is illustrated in Figure 8-7 below. There were 2 participants whose time on task fell well outside the norm: the longest of whom took 759 seconds on this task (12 minutes, 39 seconds).

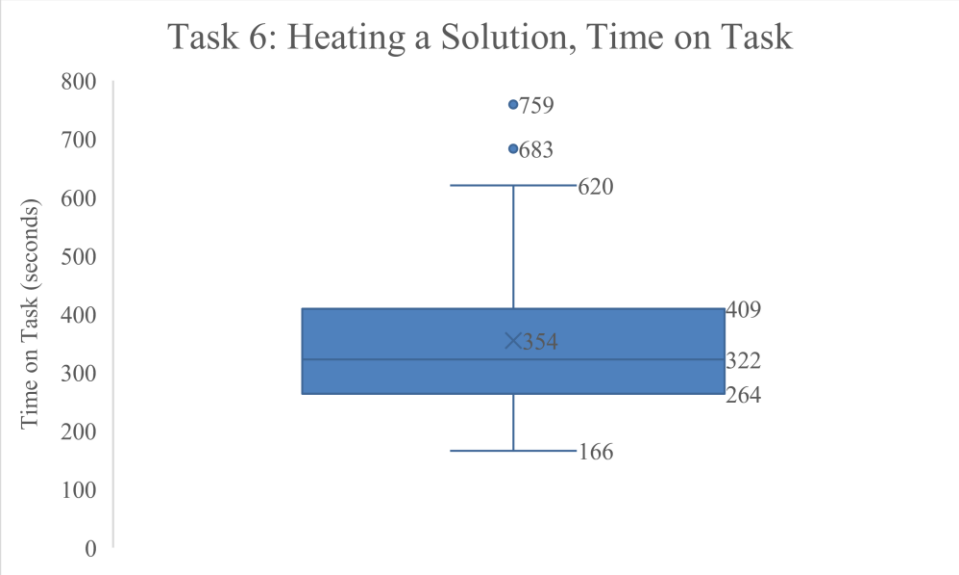

Figure 8-7: Task 6: Heating a solution, time on task (seconds) box and whisker plot

Most participants finished the task between 166 – 620 seconds, with half completing the task between 264 – 409 seconds (4.5 – 7 minutes). The average time on this task was 354 seconds, or 5 minutes 54 seconds. This was the third longest task, behind dilution and titration, respectively.

Discussion of Transcript Information from Interviews

As with other tasks, the eleven most frequently used terms are illustrated in a histogram in Figure 8-8, below. Those eleven terms are, alphabetically: Celsius, degrees, going, heat, hot, just, like, now, plate, and temperature. The most frequently used term was, again, going, with 149 occurrences, followed by just and degrees, with 110, 97 occurrences respectively.

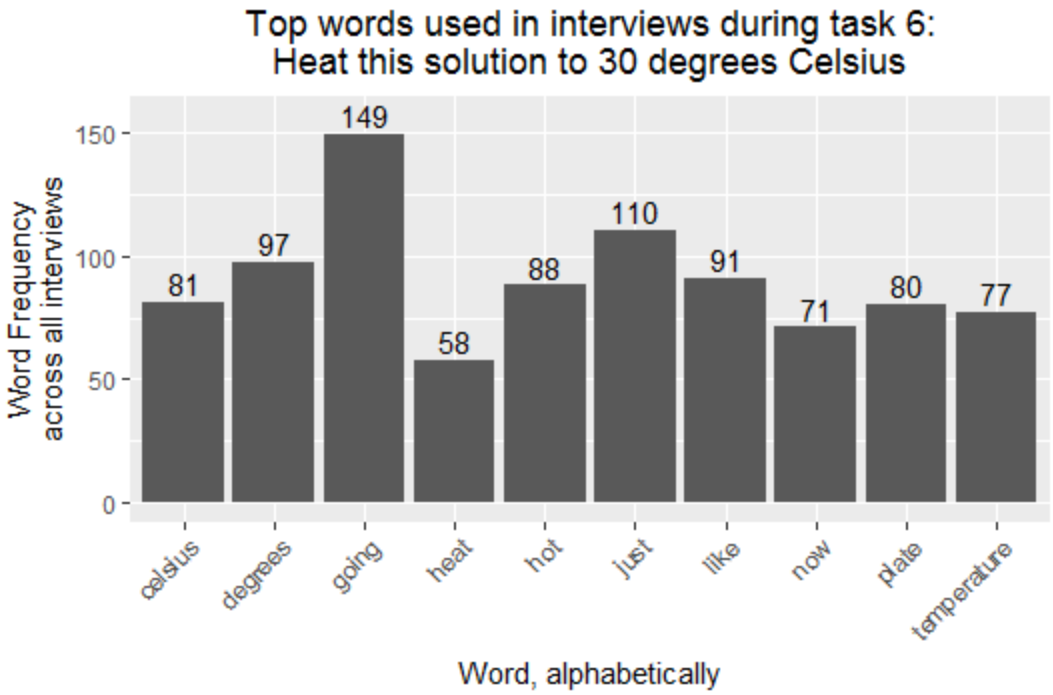

Figure 8-8: 11 most frequently used terms in Task 6

With each of these most frequent terms in such high numbers, it may seem like participants were more chatty than other tasks. However, that does not appear to be the case, when word density is considered: while there were 643 distinct terms, occurring 3343 times, the average time on task was longer also. This averaged to about nine words per second, across all participants. Again, this seems like a lot, but the previous task, transferring solution, averaged about 7 words per second. This slight uptick could also have been chatter to fill the time, or a consequence of the researcher trying to ask planning questions of students who set the hot plate to very low temperatures (causing their heating to take a very long time).

Transcripts' most frequent terms were also analyzed through a broader lens by creating a wordcloud of the 50 most frequently occurring terms during the task, Figure 8-9.

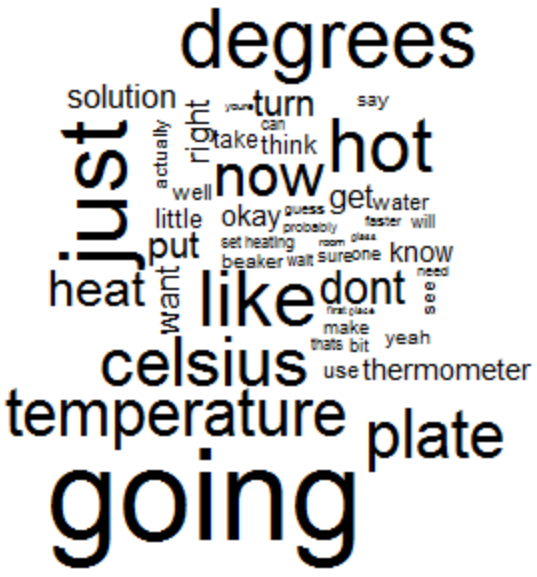

Figure 8-9: Wordcloud of the 50 most commonly used terms in Task 6

A close analysis of this wordcloud shows that it also includes terms like heating/heat/hot, beaker, thermometer, wait, water, etc. These terms' appearance on the cloud indicates that participants were, as in other tasks, interested in specific scientific terms, rather than general vague ones. There were also again filler words: little, actually, right, probably, okay, guess, wait, sure, bit. While students were waiting for their solutions to heat their growing impatience was clear by their tendency to turn up the heat setting on their hot plate.

Conclusions from Task 6

While this task did tend to take longer than expected, it was not the longest task during the interviews. The interviewer expected, based on anecdotal evidence from previous years, for a non-negligible portion of the population to select a stir plate rather

than a hot plate to heat their solutions. The undergraduate laboratories are equipped with combination stir/hot plates, so this confusion was eliminated, and further analysis on the topic is not necessary.

A finding from this study that should be communicated clearly to TAs or other instructors who interact with general chemistry students frequently is that one doesn't need to be overly concerned with over-heating solutions. Many students set their hot plates to a low temperature (below 100 ^{o}C) to heat it to 30/35^{o}C – which will get their solutions heated up, eventually, but will take a long time. Participants didn't seem to understand that by turning their hot plates higher, they would simply heat their solutions faster, and that they could simply remove their solutions when they reached or approached the desired temperature. This is a fundamental disconnect which should be addressed, and perhaps it is, but since thermodynamics are not generally taught until near the end of general chemistry, perhaps it should be addressed during the first lab in which heating a substance occurs. There is a lab at the end of the 1212 course in which they assemble a miniature solar panel, which requires annealing their titanium dioxide substrate for the panel on a hot plate. During this lab, they are instructed not to turn the hot plate too hot, for fear of causing cracks in their substrate. Perhaps that lab experience is coloring their perception of how high to set their hot plate, but that is the very last lab in the general chemistry series, so only about half of the interview participants could even have been exposed to that particular lab already before participating in this interview.

9 ANALYSIS OF TASK 7: DECANTING A SOLUTION

Introduction to the Task

The prompt for task 7 read: "Decant this solution from Beaker A to Beaker B." According to the instructions for decanting a liquid from their 1211 lab manual in lab 3: students should hold a stir rod perpendicular to and at the beaker's lip, and pour slowly and carefully until the majority of the solution has been separated from the solid.

Table 9-1: Rubric for Decanting a Solution

Exemplary (5)	• Uses glass rod across beaker top, normal to lip of beaker • Pours slowly, after solid has settled • Doesn't stir before decanting • Doesn't try to use filter paper • Slows pour as approaching end of liquid • Doesn't try to heat
Acceptable (4)	• Doesn't use stir rod • Pours slowly and gets minimal sand in beaker
Neutral (3)	• Uses rod, but uses pipet to get last bit of liquid OR • Same speed of pour throughout OR • Tries to use paper/heat
Poor (2)	• Doesn't use rod and uses pipet • Same speed throughout • Tries to use paper/heat
Very Poor (1)	• Doesn't use stir rod • Uses pipet to get last bit of liquid • Tries to use filter paper or heat • Stir/disturb solid before decanting

Discussion of Interview and Survey Statistics

During the pre-interview survey, 19 total students out of 38, or 50%, perceived decanting a liquid as easy or very easy, compared with the other 50% thinking that it was

neutral or difficult (M = 3.43, SE = 0.12). No student surveyed and interviewed thought this task was very difficult. During the post interview, however, students overwhelmingly changed their minds to make this task's most common response to neutral, but the number of very easy responses also increased, while the number of difficult responses diminished completely. A total of only 16 respondents thought the task was easy or very easy, while 19 thought the task was neutral – neither easy nor difficult. These responses can be viewed via histogram in

Figure 9-1 & Figure 9-2, below.

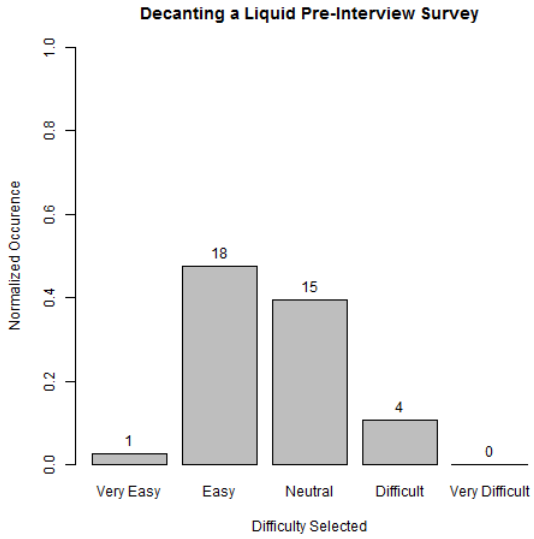

Figure 9-1: Pre-Interview Score, Decanting a Liquid

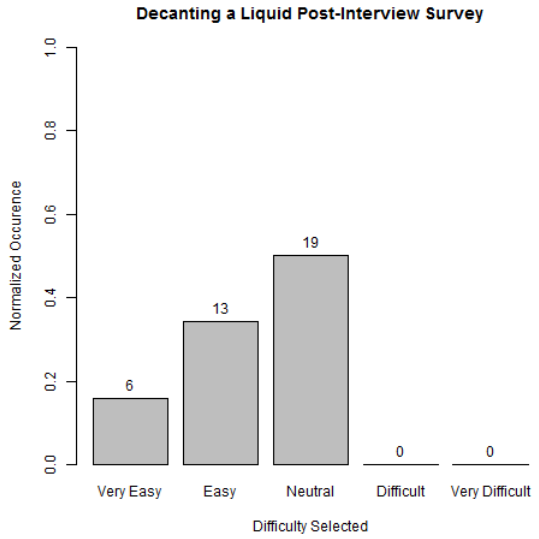

Figure 9-2: Post-Interview Survey, Decanting a Liquid

Interestingly, this shift from modal response of easy to modal response of neutral did not significantly affect the average score of perceived difficulty, (M = 3.63, SE = 0.13). This was a non-significant change: -0.200, t(34) = -1.313, p = .198. This shift must have been non-significant because of the simultaneous increase in both very easy and

neutral responses, paired with a decrease in easy responses. There was no significant Pearson correlation between pre- and post-interview responses for this item, r (35) = 0.250, p = .147. This was one of the sole survey items for which the pre- and post-interview scores were not correlated. There is no contradiction here, however. The Pearson test coming back that there was not a strong correlation between pre- and post-interview scores of perceived difficulties simply suggests that the shape of the responses to the questions do not mirror one another in the pre- and post-surveys. This is easily shown by looking at those responses in a histogram, as in Figure 9-1 and Figure 9-2, above. While in other tasks, these histograms have been shaped similarly in pre- and post-interview scores, for this task that simply was not the case.

In Figure 9-3, below, we see the distribution of scores on the interview task, (*M* (38) = 2.66, *SD* = 1.07). Both the mean and modal score are neutral in this case, with 21 interviewees receiving a neutral score.

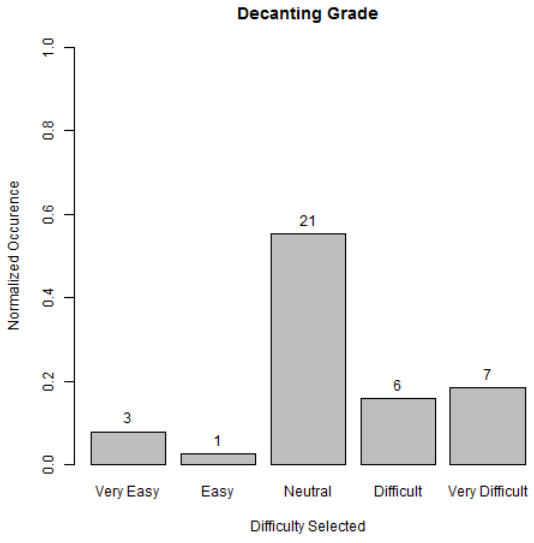

Figure 9-3: Interview Scores, Task 7 Decanting

While there is no significant difference in pre- and post- interview survey scores, there is a significant difference from both to the grades assigned from the rubric. This resulted from the rubric requiring correct use of a stir rod in the decanting process to get full marks on this task. This discrepancy calls for a reconsideration of the rubric itself, to see whether that rubric is fully representative of the goals of the teaching body for students to accomplish when completing this task.

Discussion of Common Interview Behavior

Of the 38 students who completed the decanting task, 6 disrupted the solid, 7 poured fast (though this was entirely subjective, there was no scale to differentiate fast from slow), and five used a transfer pipet to remove the last bit of liquid.

The five who used a transfer pipet automatically had 2 of the five possible points for the task deducted from their score: the purpose of decanting, according to their lab manual, is to separate as much of the solution from the solid as possible, not to fully dry the solid or recover all of the liquid. In interview questions, while not specifically probed on this action, there seemed to be a focus on recovery of the solid (as opposed to the solution) for further use. Interestingly, there was a lab in their first semester (the copper cycle lab) where they decanted to separate and were explicitly isolating the solution at one point during a decanting step (in addition to the solid) – so they should have been familiar with solution recovery from decanting as well as solid recovery, although they didn't seem to be.

Only 6 total used a stirring rod, and of those six, 3 used the stirring rod incorrectly. Six stated that the solid needed to be heated after decanting the liquid off. A

different six said that the liquid needed to be filtered to remove excess solid remaining in the liquid. Of those 6, four (67%) actually moved to begin the filtration (gravity filtration) process. None were allowed to complete the filtration process, since the task was to decant the liquid from the solid and neither filtration nor heating solid to dryness is a part of that task as performed correctly.

There was an overall lack of concern for touching the solutions (all of the solutions) in this interview process, but nowhere was this more apparent than during this task. Students seemed to be utterly unconcerned with touching their skin to the solutions during the decanting process. While during this interview, this was distilled water and sand, in a lab, which this was supposed to model, this solution could have been a corrosive acid. One has to wonder if these students are any more careful when they are fully briefed on the dangers of the chemicals they are dealing with, or whether they research the MSDS of these chemicals (outside of their TA trainings at the beginning of each lab) in case of more dangerous reagents.

Discussion of Transcript Information from Interviews

In Figure 9-4, below, we see a histogram of most frequently used terms in the seventh task of the interviews, decanting a solution from a solid in a beaker. These eleven terms were, alphabetically: beaker, don't, filter, get, going, just, like, liquid, much, pour, solid. The most frequently used term was, unsurprisingly, solid. Other scientific terms were liquid, beaker, pour, and filter. The remaining frequent terms were split between filler and action words, don't, get, going were actions while just, like, and much were fillers.

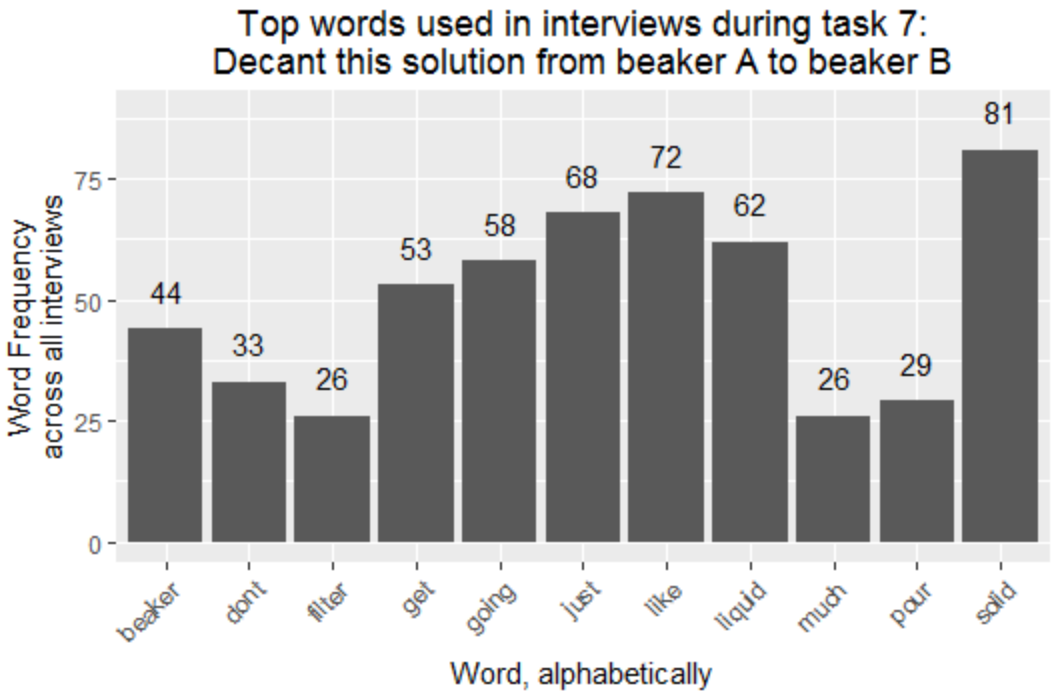

Figure 9-4: 11 Most frequently used terms in Task 7

In the wordcloud of the fifty most frequently used terms of the seventh task in the interview (Figure 9-5 below), you can see that scientific/naming words seemed to dominate the picture: in addition to the most frequent terms, you also see water, sand decanting, decant, and pipet. However, filler words also have a strong presence in the image: guess, something, okay, know, really, usually, kind, sure, think, try, probably.

While action words are certainly there, they are much less prevalent: put, use, can, and make are all substantially smaller than most other words in the image.

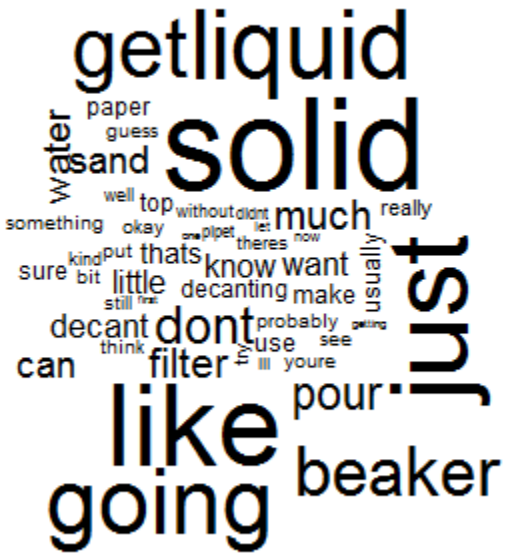

Figure 9-5: Wordcloud of 50 most frequently used terms in Task 7

Conclusions from Task 7

Given how very few students actually used a stir rod to complete their decanting task, it is worth re-examining the motivations for using a stir rod while decanting – and if they are thin, it may be worth considering removing that instruction, or specifying that this instruction is specific to that experiment and not for decanting as a whole.

10 ANALYSIS OF TASK 8: TITRATION WITH A VISUAL INDICATOR

Introduction to the Task

The eighth and final task for all students eight tasks were offered to was a titration with a visual indicator, methyl orange. This titration was with 1M NH_3 and 1M HCl. The prompt read:

> "Given this 1.0 M HCl, 1.0 M NH_3, and Methyl Orange as a visual indicator, please titrate 10.00 mL of NH_3 with HCl. The methyl orange turns from a tangerine color to a pale rose gold at the end point, and is a bright pink past the end point."

There was not a prompt in the 1211 manual for a simple acid-base titration with a visual indicator, without performance of a back titration, hence writing this prompt in the style of the manual.

The rubric, which follows in Table 10-1 on the following page, for performance of a titration with a visual indicator, still follows from the instructions to perform a titration laid out in the first version of the Norton manual. Special attention was paid to which reagents were used within their manual.

Table 10-1: Rubric for Titration with a Visual Indicator

Exemplary (5)	• Uses Erlenmeyer flask to perform titration • Indicator input to solution as primary step • Larger, ~0.25 mL drops from buret near beginning, smaller, tiny droplets nearer endpoint as change begins to occur • Touches side of buret tip to side of glassware • Swirls after each individual droplet, or uses stir bar & plate • Writes color changes as they occur • Stops at correct endpoint • Records beginning and final volumes in buret
Acceptable (4)	One of: • Does not swirl/use stir bar & plate • Does not write observations • Does not slow as color endpoint approaches • Uses beaker, rather than Erlenmeyer flask
Neutral (3)	One of: • Switches solution and titrant • Forgets indicator, but sees that indicator was omitted and re-completes • Goes <1 mL past end point Or two of: • Does not swirl/use stir bar & plate • Does not write observations • Does not slow as color endpoint approaches • Uses beaker, rather than Erlenmeyer flask
Poor (2)	Three of: • Switches solution and titrant • Forgets indicator, but sees that indicator was omitted and re-completes • Goes <1 mL past end point • Does not swirl/use stir bar & plate • Does not write observations • Does not slow as color endpoint approaches • Uses beaker, rather than Erlenmeyer flask
Very Poor (1)	Four or more of: • Switches solution and titrant • Forgets indicator, but sees that indicator was omitted and re-completes • Goes <1 mL past end point • Does not swirl/use stir bar & plate • Does not write observations • Does not slow as color endpoint approaches • Uses beaker, rather than Erlenmeyer flask

Discussion of Interview and Survey Statistics

Students perceived difficulty of a titration with a visual indicator are illustrated in Figure 10-1, Figure 10-2, below.

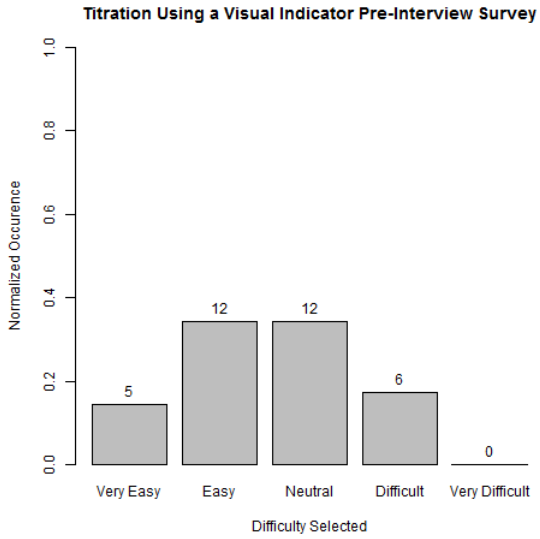

Figure 10-1: Pre-Interview Survey Scores of Titration with a Visual Indicator

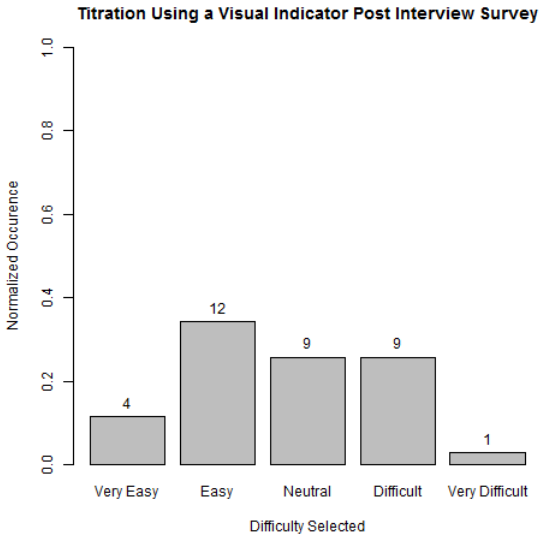

Figure 10-2: Post-Interview Survey Scores of Titration with a Visual Indicator

In the pre-interview survey, which was offered to all participants, 17 responded that performing a titration using a visual indicator was easy or very easy, while a total of 35 answered the question, (M = 3.44, SE = 0.17), while in the post-interview survey, a total of 35 respondents answered, with 16 of those 35 responding that the task was easy or very easy (M = 3.25, SE = 0.19), for an overall score of both tasks in the pre- and post-surveys of easy. This did not represent a significant change in perceived difficulty of the task, $t(31) = 1.10$, $p = .280$.

Students' performance is illustrated in Figure 10-3, on page 89. Overall, students scored a 1, with 12 of the 20 to whom it was offered receiving a 1. Again, this discrepancy in perceived difficulty versus performance is more likely a denunciation of the rubric than of students' lack of self-awareness.

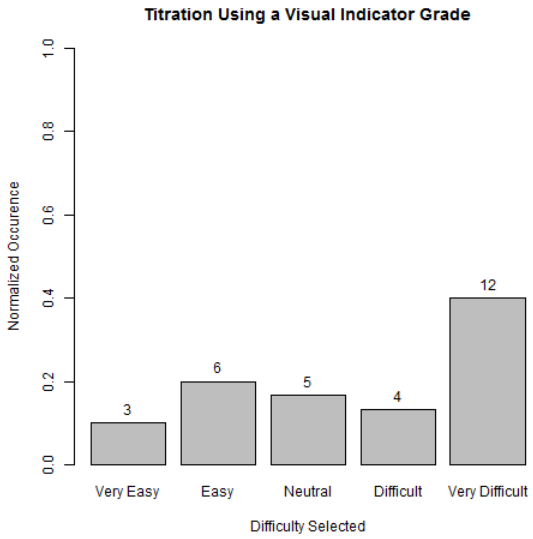

Figure 10-3: Titration with a Visual Indicator Score

This represents a significantly different score than the overall score on the task, which fell squarely into the neutral range (M = 2.47, SD = 1.46).

Discussion of Common Interview Behavior

This task gave nearly every student it was presented to a great deal of trouble. The trouble ranged from not being told which reagent was the titrant, to being unclear on where to put the indicator, to using the incorrect glassware for a titration in general. Some students, when asked directly during the interview portion (after the observations portion) stated that they did not know what a titration was, or that they did not know the purpose of a titration. Several stated that they did know what one was and that they had performed it incorrectly (and knew it) but did not know how to perform it correctly.

For an explicit count of issues from the observation of this task we have the following:

This task was offered to 28 of the 38 participants, or was not offered to 10 participants. Of the 28 to whom this task was offered, there were several who had glassware errors: 12 did not use an Erlenmeyer flask, but used a beaker instead (43%). 3 did not use a buret for their titration: they used a combination of their beakers and graduated cylinders to perform a variation on a titration (11%). 4 did not perform a titration at all – they moved around solutions, tried to figure out what to do, and ultimately gave up rather than finishing the task (14%). While this may not be considered a glassware error, 5/28 students reversed the titrant and the (receptor) – so they put the ammonia into the buret while putting the HCl into their titration receptacle. Interestingly, none of the five who did this stopped when their solution was already bright pink on addition of the methyl orange, which should have been an early warning that they had reversed which reactant went in which position. This is an indictment of students' understanding not only of titrations themselves, but of what to look for in titrations.

There were also several students who either did not record applicable volumes, or who were prompted to record the volumes which they recorded: 11 did not record the initial volume, but recorded the final volume. A different 11 did not record the final or initial volumes. 9 of these did-not-record students recorded neither initial nor final volume, the 2 remaining for each (final, initial) were distinct four students. 1 student recorded the final volume after prompting. Related to the students who did not record volumes: 4 students (14%) did record their observations, which, given that this task was largely qualitative in measure, was exceptionally small. Finally, 10 of the 28 students given the task to perform the titration went past the end point – well into bright pink territory.

Most of the information gathered from the titration task was qualitative, as well as the task itself being qualitative: Students' quotes regarding the titration were often more academically useful to this study than their performance of the task itself. These quotes informed the research team of students' thought processes while completing titrations.

Discussion of Transcript Information from Interviews

Task 8: titration with a visual indicator interview transcripts were analyzed with the same lens as all other tasks in the interview. Transcripts were separated by task and then using R's tm, wordcloud, and ggplot2 packages were mined for descriptive information. English stopwords were removed from the analyses for the graphics below, transcripts were stripped of punctuation and numbers as well as whitespace before these graphics were generated.

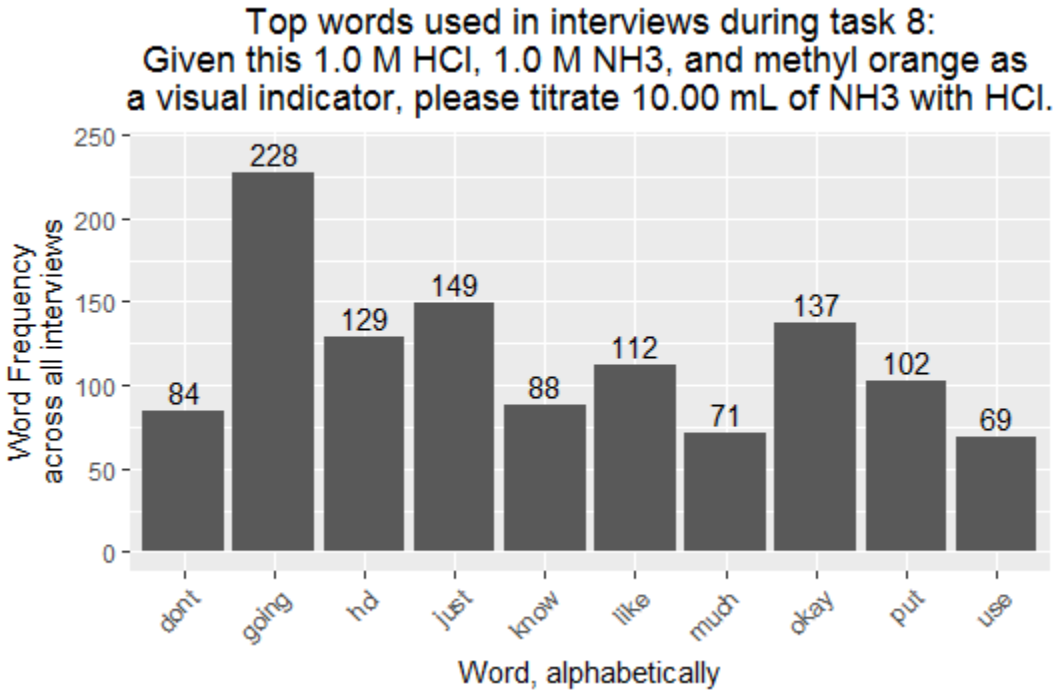

Figure 10-4: 10 most frequently used in Task 8 across all participants

In Figure 10-4, above, we see a histogram depicting term frequency of the 10 most commonly occurring phrases during task 8 in these interviews. These 10 terms were (in alphabetical order) don't, going, HCl, just, know, like, much, okay, put, and use.

Going was the most frequently used term with 228 occurrences across 28 interviewees who participated in this interview task. That comes out to about 8 uses of going per participant, in this task. This indicates that participants were more comfortable (whether consciously or subconsciously) with describing what they were doing, either before or while they were doing these things.

Filler words just, like, and okay also appeared in the 10 most frequently occurring terms, with 149, 112, and 137 occurrences, accordingly. That makes up (398/1169) of these top terms, or 34% of those top term occurrences. In this situation only, know was not included in filler words, specifically because don't occurred 84 times, to know's 88.

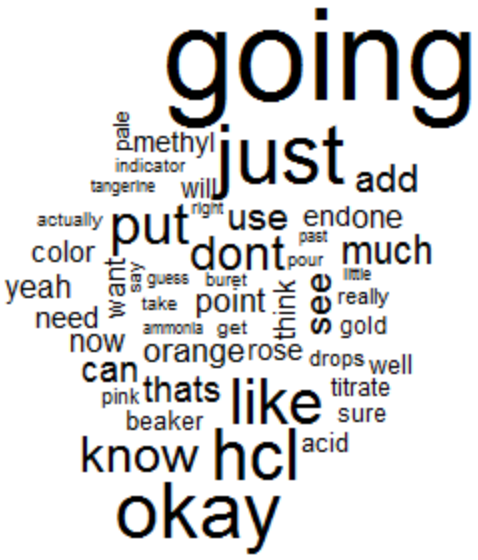

Figure 10-5: Word cloud of 50 most frequently used words during titration task

In Figure 10-5, above, we see a wordcloud of the 50 most frequently occurring terms during task 8 in these interviews.

Interestingly, while participants mentioned methyl orange by name, mentioned the acid by name, did not nearly as often mention the base, NH_3, by name. While not particularly obvious from observation of the 10 most frequently used terms, an examination of the wordcloud sees very frequent use of color words: pink, tangerine, orange, rose, gold. These could be considered scientific descriptors, especially since they were specifically mentioned in the prompt, which was removed from these analyses of the transcripts from the task.

11 ANALYSIS OF TRANSCRIPTS FROM ALL TASKS IN INTERVIEWS

Transcripts were analyzed by word frequency counts, overall, by sex, and by class. In all analyses, the most used word was going.

Words used most frequently can be classified as action words, planning words, naming words, and filler words. Examples of each are included in the table below, Table 11-1:

Table 11-1: Sorting convention of example words from transcripts

Action Words	Planning Words	Naming Words	Filler Words
• Going • Can • Use • Put • Get	• Know • Don't • Now • Think • Want • Need	• Beaker • Solution • Cylinder • Water • Flask	• Okay • Just • Like • Yeah • Um • Mhm

Often these frequently used words are used in conjunction with one another to form a phrase that would fall into a different overall category than the word itself:

"I think I can use this cylinder to transfer the solution to the beaker"

"I know I need 100 mL of solution at the end of this"

"Okay, so if I can get on eye level, I will see the actual reading instead of like, an overestimation"

The most used word overall and by every sub-analysis was "going". Going had a frequency across all interviews of 963. Of those 963, 595 instances (61.8%) were from female students. This reflected the overall sex distribution of the interviewees (68.4% female). Interestingly, despite there being 52.6% of the interview population in 1211

class, the word going was more commonly used by 1212 students (42.3% of the instances of going were by 1211 students). This could indicate that Chemistry 1212 students were more comfortable with the part of the interview in which they said aloud what they were about to do (or what they were already doing) rather than what they had just done.

Another word among the top ten most used in each sub-analysis was “just”. Just had an overall frequency of 658 occurrences, making in the second most used word throughout all interviews. It was the second most used word among both male and 1211 interviewees, third amongst 1212 interviewees, and fourth amongst female interviewees. Interestingly, although “just” was only the fourth most popular word used by female interview participants, those participants accounted for 79.9% of its use. This is significantly larger than the percentage of interview participants who were female, and represents an uptick in usage by female participants over male participants.

The third most frequently used word overall was “okay”. Okay was second most frequent amongst 1212 and female sub-groups, third most frequent with the 1211 subgroup, and fourth most frequent with male interviewees. Female participants accounted for 76.3% of use of the word okay, while 1212 students accounted for 54.3% of its use.

The most frequently used naming word was beaker, overall, followed by solution. Beaker was used 331 times throughout all interviews, with 1212 students using it more than 1211 (56.7% of occurrences) and female students using it more than male (65.6% of occurrences). These distributions of use of the word beaker were well aligned with the interview population, which was 54.6% 1211 students, and 68.4% female. Solution occurred 304 times overall, with 162 of those occurrences in 1211 interviews (56.1%)

and 172 of them being from female participants (59.5%). This indicated that male participants used the word solution a bit more than female participants, when per capita use is considered.

From 020203:

> "Okay so, I'm going to get down on eye level and I'm going to look at where the little um meniscus is, the little dippy thing, and it looks like it's exactly on the 30 1, 2, 3, 32 line, it might even be a little bit above that so I'm going to do 32.1 mL."

Within each task, there were several instances of participants who were chatty, as well as of participants who said very little. The above quote was from a particularly wordy participant, they said nearly 5000 words in the interview (average = 2094 +/- 1050, median = 1851, mode = 2717). While this participant did use relaxed language including several filler words, they also used a naming word (meniscus) and it was clear that they were making efforts to make their thoughts and actions well understood by anyone watching this footage at a later date.

12 FINAL CONCLUSIONS

Should this research be conducted again, or carried out to draw further conclusions from it, all inferences from the rubrics should be drawn only under the conditions that the rubrics are reassessed for validity. This would be completed via close study of inter-rater reliability and discussion of whether the rubric would make more sense if drawn on more than this single set of manuals. A general, portable validated rubric should be created, and possibly has been created, simply not published. This rubric could be a stepping stone towards more reliable national portability of chemistry education through the general chemistry curriculum.

One step towards that more reliable, viable rubric was to conduct the survey conducted in December 2016 – to ask students (who are not being asked to perform the task, and are not being recorded) to select the glassware they would use if asked to perform the task. In that survey, they were also given the option to select glassware/equipment not in the list via a free text option. In this way, a larger sample of students' anticipated response to tasks can be gathered. From this it can be determined whether the glassware penalties on the current rubrics were appropriately discriminatory for task performance, or if students' knowledge from their manuals was truly insufficient and this is the reason for their performance.

In each task performed during the interview process, students' performance as measured by the rubrics laid out at the beginning of this study did not closely match their perceived difficulty of the skills measured in those interview tasks. This mismatch in

perceived difficulty versus performance persisted regardless of participant sex or class level.

REFERENCES

1. Bauer, C. F., Attitude towards Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impacts. *Journal of Chemical Education* **2008,** *85* (10), 1440 - 1445.
2. Buck, L. B.; Bretz, S. L.; Towns, M. H., Characterizing the Level of Inquiry in the Undergraduate Laboratory. *Journal of College Science Teaching* **2008**.
3. Cacciatore, K. L.; Sevian, H., Incrementally Approaching an Inquiry Lab Curriculum: Can Changing a Single Laboratory Experiment Improve Student Performance in General Chemistry? *Journal of Chemical Education* **2009,** *86* (4), 498 - 506.
4. Fay, M. E.; Grove, N. P.; Towns, M. H.; Bretz, S. L., A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chemistry Education Research and Practice* **2007,** *8* (2), 212 - 219.
5. Galloway, K. R.; Malakpa, Z.; Bretz, S. L., Investigating Affective Experiences in the Undergraduate Chemistry Laboratory: Students' Perceptions of Control and Responsibility. *Journal of Chemical Education* **2016,** *93* (2), 227-238.
6. Towns, L. B. B. a. M. H., Preparing Students To Benefit from Inquiry-Based Activities in the Chemistry Laboratory: Guidelines and Suggestions. *Journal of Chemical Education* **2009,** *86* (7), 820 - 822.
7. Walker, J. P.; Sampson, V.; Zimmerman, C. O., Argument-Driven Inquiry: An Introduction to a New Instructional Model for Use in Undergraduate Chemistry Labs. *Journal of Chemical Education* **2011,** *88* (8), 1048-1056.
8. Weaver, G. C.; Russell, C. B.; Wink, D. J., Inquiry-based and research-based laboratory pedagogies in undergraduate science. *nature chemical biology* **2008,** *4* (10), 577 - 580.
9. Docktor, J.; Heller, K. In *Robust Assessment Instrument for Student Problem Solving*, Proceedings of the NARST 2009 Annual Meeting, Garden Grove, California, April 2009; Garden Grove, California, 2009.
10. Mueller, J. Authentic Assessment Toolbox: Rubrics. http://jfmueller.faculty.noctrl.edu/toolbox/rubrics.htm.
11. Mueller, J. Authentic Assessment Toolbox: How Do You Create Authentic Assessments? http://jfmueller.faculty.noctrl.edu/toolbox/howdoyoudoit.htm.
12. Council, N. R., *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press: 2012.
13. Abdullah, M.; Mohamed, N.; Ismail, Z. H., The effect of an individualized laboratory approach through microscale chemistry experimentation on students' understanding of chemistry concepts, motivation and attitudes. *Chem. Educ. Res. Pract.* **2009,** *10* (1), 53-61.
14. Anders, C.; Berg, R., Factors related to observed attitude change toward learning chemistry among university students. *Chem. Educ. Res. Pract.* **2005,** *6* (1), 1 - 18.

15. Barbera, J.; Adams, W. K.; Wieman, C. E.; Perkins, K. K., Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *Journal of Chemical Education* **2008,** *85* (10), 1435 - 1439.
16. Bauer, C. F., Beyond "Student Attitudes": Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning. *Journal of Chemical Education* **2005,** *82* (12), 1864 - 1871.
17. Chan, J. Y. K.; Bauer, C. F., Effect of peer-led team learning (PLTL) on student achievement, attitude, and self-concept in college general chemistry in randomized and quasi experimental designs. *Journal of Research in Science Teaching* **2015,** *52* (3), 319-346.
18. Cheung, D., Developing a Scale to Measure Students' Attitudes toward Chemistry Lessons. *International Journal of Science Education* **2009,** *31* (16), 2185-2203.
19. Coll, R. K.; Dalgety, J.; Salter, D., The Development of the Chemsitry Attitudes and Experiences Questionnaire (CAEQ). *Chemistry Education Research and Practice in Europe* **2002,** *3* (1), 19 - 32.
20. Jr., I. A. S.; Zimmaro, D. M., The Influence of Collaborative Learning on Student Attitudes and Performance in an Introductory Chemistry Laboratory. *Journal of Chemical Education* **2002,** *79* (6), 745 - 749.
21. Marsh, H. W.; O'Neill, R., Self Description Questionnaire III: The Construct Validity of Multidimensional SelfConcept Ratings by Late Adolescents. *Journal of Educational Measurement* **1984,** *21* (2), 153 - 174.
22. LOGAN, G. D., Attention and Automaticity in Stroop and Priming Tasks: Theory and Data. *COGNITIVE PSYCHOLOGY* **1980,** *12*, 523-553.
23. Ratcliff, R.; McKoon, G., A Retrieval Theory of Priming in Memory. *Psychological Review* **1988,** *95* (3), 385 - 408.
24. Tulving, E.; Hayman, C. A. G.; Macdonald, C. A., Long-lasting Perceptual Priming and Semantic Learning in Amnesia: A Case Experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **1991,** *17* (4), 595 - 617.
25. Kozma, R. B.; Russell, J., Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *JOURNAL OF RESEARCH IN SCIENCE TEACHING* **1997,** *34* (9), 949 - 968.
26. arc12 Standard Set of English Stopwords. https://github.com/arc12/Text-Mining-Weak-Signals/wiki/Standard-set-of-english-stopwords (accessed 02/21/2017).
27. Brandt, S.; Lieven, E.; Tomasello, M., German Children's Use of Word Order and Case Marking to Interpret Simple and Complex Sentences: Testing Differences Between Constructions and Lexical Items. *Lang Learn Dev* **2016,** *12* (2), 156-182.
28. Droge, A.; Fleischer, J.; Schlesewsky, M.; Bornkessel-Schlesewsky, I., Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: Electrophysiological evidence from non-canonical linearizations in a flexible word order language. *Brain Res* **2016,** *1633*, 149-66.
29. Sung, J. E.; Yoo, J. K.; Lee, S. E.; Eom, B., Effects of age, working memory, and word order on passive-sentence comprehension: evidence from a verb-final language. *Int Psychogeriatr* **2017**, 1-10.

APPENDICES

A RUBRICS FOR EACH SCIENTIFIC SKILL

Selecting Proper Glassware	
Exemplary (5)	Selects volumetric flask for dilution. Selects volumetric pipet for transfer 10.00 mL question. Uses funnel to fill buret for titration. Uses stir rod for decanting. Appropriate sized glassware for titration.
Acceptable (4)	Misses one selection from exemplary
Neutral (3)	Misses 2 selections from exemplary
Poor (2)	Misses 3 selections from exemplary
Very Poor (1)	Misses 4 or more selections from exemplary

Recording Data	
Exemplary (5)	Records volume in graduated cylinder task 1 Records volume of NaCl transferred in task 2 Records mass in task 3 Records initial & final buret readings in task 5 Subtracts initial & final buret readings in task 5 to confirm amount transferred Records initial & final temperatures in task 6 Records amount of NH_3 transferred to receptacle in task 8 Records initial & final buret readings of HCl in task 8
Acceptable (4)	Misses 1 – 2 from exemplary
Neutral (3)	Misses 3 – 4 from exemplary
Poor (2)	Misses 5 – 6 from exemplary
Very Poor (1)	Misses 7 or more from exemplary

Recording Observations	
Exemplary (5)	Records identity of solid in task 3 Records approximate time* in task 6 Records initial color of indicator in task 8 Records color on addition to reactant flask/beaker in task 8 Records color at end point of titration in task 8 Records color past end point in task 8 (if applicable)
Acceptable (4)	Misses task 3 or 6 recordings, but records at least one color in task 8
Neutral (3)	Misses task 3 & 6 recordings but records at least one color in task 8
Poor (2)	Misses task 3 or 6 recordings and does not record observations in task 8
Very Poor (1)	Does not record any observations throughout 8 task interview

Reading a Graduated Cylinder (Task 1)	
Exemplary (5)	Reads at eye level. Uses correct number of significant figures Records value with units Value is within +/- 0.2 mL of researcher's recorded value

Acceptable (4)	Within +/- 0.3 mL
	One of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Neutral (3)	Within +/- 0.5 mL
	Two of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Poor (2)	Outside +/- 0.5 mL
	Three of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Very Poor (1)	Outside +/- 0.5 mL
	Four or more of the following:
	• Prompted to record
	• Missing one:
	  ○ Significant Figures
	  ○ Units
	• Reads near but clearly not at eye level
	• Reads from elsewhere than bottom of meniscus

Use of a Volumetric Flask	
Exemplary (5)	uses volumetric pipet to place 20.00 mL NaCl into flask initially. Dilutes with H_2O to line. Caps and inverts after adding some but not all of water. Slows pouring at narrowing of neck. Uses transfer pip to add water dropwise to line. Caps and inverts again to ensure proper mixing. Uses funnel.
Acceptable (4)	Does not use funnel to pour and/or does not use transfer pipet for last bit to line Still doesn't fill past line OR Uses large graduated cylinder (loses decimal place) to put 20 mL but otherwise correct
Neutral (3)	Adds 20.00 mL to flask initially, generally uses properly but either Does not cap + invert Fills (small amount) past line
Poor (2)	fills past line and does not invert. Does not use funnel, does not use transfer pipet.
Very Poor (1)	Doesn't use line. Still uses 100 mL dilution. OR

Performance of Dilution Task	
Exemplary (5)	Use volumetric flask without prompt Perform calculations correctly Use volumetric pipet twice Fill to neck with funnel/pour, then use smaller pour/transfer pipet to finish fill to etch
Acceptable (4)	
Neutral (3)	Use graduated cylinder/other glassware but dilution is perfect
Poor (2)	
Very Poor (1)	Does not complete dilution Uses beakers as exact measures of volume Uses volumetric flask (200 mL) to perform 10/100 mL dilution

Weighing a solid	
Exemplary (5)	Cleans first uses weigh paper/boat + scoopula proper tare closes door waits for stable value records all digits + unit
Acceptable (4)	doesn't brush clean doesn't wait for stable value All other from exemplary
Neutral (3)	doesn't close door, or leaves off digit/unit, or both doesn't brush and doesn't wait improper tare
Poor (2)	two from neutral
Very Poor (1)	three or more from neutral

Use of Volumetric Pipet	
Exemplary (5)	Is using volumetric pipet Checks fit of pipet in holder Tests with small amount of liquid Slows at wide part
Acceptable (4)	Uses volumetric pipet but doesn’t check fit or test small amount
Neutral (3)	Is using graduated but uses correctly Doesn't slow at wide or overshoots line but gets back down to line
Poor (2)	Uses graduated at 4 level or Uses volumetric and doesn’t check fit/test small amount and doesn’t slow
Very Poor (1)	Uses graduated but still wrong Overshoots/undershoots

Use of Mohr Pipet	
Exemplary (5)	Checks fit of pipet in holder Tests with small amount of liquid Slows as approaches 10, 0 mL lines Stops at 10/0 mL line at bottom of pipet
Acceptable (4)	Stops at 0 mL line and one of: Overshoots 10 mL line at top of pipet but rolls back down Doesn't test Doesn't check fit
Neutral (3)	Stops at 0 mL line and two of: Overshoots 10 mL line at top of pipet Doesn't test Doesn't check fit
Poor (2)	Goes past 10/0 line but expresses that shouldn't have
Very Poor (1)	Goes past 10/0 line with no address

Performance of Task 4: Transfer 10.00 mL of solution	
Exemplary (5)	Uses Mohr or volumetric pipet unprompted at 5 level
Acceptable (4)	Uses Mohr or volumetric pipet unprompted at 4 level
Neutral (3)	Use of pipet unprompted at 3 level
Poor (2)	Use of pipet unprompted at 2 level
Very Poor (1)	Use of pipet unprompted at 1 level
Requiring a prompt to complete this task via pipet (first completing via graduated cylinder etc.) is a 1 point deduction from performance, and is dependent on use of other glassware	

Reading a Buret	
Exemplary (5)	Glass height adjusted for student Numbers toward student Reads from bottom of meniscus Records volume to +/- 0.03 mL
Acceptable (4)	Missing one from exemplary
Neutral (3)	Volume to +/- 0.07 mL doesn't record unit
Poor (2)	Missing 2 from exemplary Both from neutral
Very Poor (1)	Volume more than .07 mL off and does not record unit. Does not read from eye level. Glass not height adjusted.

Using a Buret	
Exemplary (5)	Reads properly. Turns stopcock slowly, only allows a drip not a full stream. Slows speed near final volume dispensed. Ensures stopcock closed before filling.
Acceptable (4)	Can't read properly because of height Full stream at first but slows before anticipated point
Neutral (3)	Two of: stopcock open to fill but otherwise 5 use. allows stream but still stops at anticipated point. 3 use of reading, but otherwise 5 use of buret. over/undershoots anticipated volume dispensed
Poor (2)	Neutral use AND Height adjustment OR Reading at 3 level AND acceptable
Very Poor (1)	Combination of 2 or more from neutral. OR reading at 1 level.

Using a Thermometer	
Exemplary (5)	Thermometer lifted off bottom of glass but fully within solution Read from eyelevel and normal to gaze Held in solution long enough to read temperature
Acceptable (4)	Not at eye level or Not normal to gaze
Neutral (3)	Thermometer in liquid but touching bottom glass or Not in substance long enough to tell temperature
Poor (2)	One each from 3 & 4
Very Poor (1)	Thermometer touches glass Thermometer not held in solution long enough to equilibrate Not at eye level Not normal to gaze

Using a Hot Plate	
Exemplary (5)	Selects proper equipment. Plugs in before turning on. Does not turn on empty. After reaching temperature switches plate off. Cools/unplugs before putting away
Acceptable (4)	One of: Turns on empty Doesn't turn off plate after reaching temperature Doesn't cool off/unplug before putting away
Neutral (3)	Starts with stir plate but completes rest from 5 OR Doesn't plug in before heating but then completes after realizing mistake
Poor (2)	Both of 3 and one of 4
Very Poor (1)	2 or more of items from 3 or 4.

Performance of Task 6: Heating a Solution	
Exemplary (5)	5 use of thermometer and hot plate
Acceptable (4)	4 use of one
Neutral (3)	4 use of both or 3 use of one
Poor (2)	3 use of both or 2 use of one
Very Poor (1)	2 or worse use of both

Decanting a solution	
Exemplary (5)	Uses glass rod across beaker top, normal to lip of beaker Pours slowly, after solid has settled Doesn't stir before decanting Doesn't try to use filter paper Slows pour as approaching end of liquid Doesn't try to heat
Acceptable (4)	Doesn't use stir rod Pours slowly and gets minimal sand in beaker
Neutral (3)	Uses rod, but uses pipet to get last bit of liquid OR Same speed of pour throughout OR Tries to use paper/heat
Poor (2)	Doesn't use rod and uses pipet Same speed throughout Tries to use paper/heat
Very Poor (1)	Doesn't use stir rod Uses pipet to get last bit of liquid Tries to use filter paper or heat Stir/disturb solid before decanting

B LIST OF MATERIALS AVAILABLE TO INTERVIEW PARTICIPANTS

Sugar
NaCl
1M NH_3
1M HCl
Distilled water
8 250-mL beakers
2 150-mL beakers
2 50-mL graduated cylinders
2 50-mL burets
2 10-mL graduated cylinders
2 10-mL Mohr (graduated) pipets
2 10-mL volumetric pipets
2 pipet pumps
Transfer pipets
2 scoopulas
2 sets of tweezers
Beaker tongs
Crucible tongs
Wire mesh
2 thermometers
2 glass stirring rods
1 magnetic stir bar
1 magnetic stir plate
1 hot plate, with ^{o}C increment markings
1 analytical balance
2 pair goggles
2 funnels
Weigh boats
Weigh paper
Filter paper