PHYSICS COMMUNICATION AND PEER ASSESSMENT IN A REFORMED INTRODUCTORY MECHANICS CLASSROOM

A Dissertation
Presented to
The Academic Faculty

By

Scott S. Douglas

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology

August 2018

PHYSICS COMMUNICATION AND PEER ASSESSMENT IN A REFORMED INTRODUCTORY MECHANICS CLASSROOM

Approved by:

Dr. Michael Schatz, Advisor
School of Physics
Georgia Institute of Technology

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Dr. Flavio Fenton
School of Physics
Georgia Institute of Technology

Dr. Wendy Newstetter
College of Engineering
Georgia Institute of Technology

Dr. Edwin Greco
School of Physics
Georgia Institute of Technology

Date Approved: May 25, 2018

On two occasions I have been asked, — “Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?” ...I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

Charles Babbage

[We] should not assume that experts are the only ones who have a right to express themselves... I am very conscious of the fact that our feelings and strivings are often contradictory and obscure and that they cannot be expressed in easy and simple formulas.

Albert Einstein

This thesis is dedicated to Anat and to the new person who will shortly join our family. The former I have loved for many years, together and apart, and has taught me innumerable things—the latter, I expect, will teach me even more.

ACKNOWLEDGEMENTS

I would like to thank Mike Schatz and Ed Greco for their continual support and encouragement. The introductory physics program they helped to build at Georgia Tech has been the testing-ground for all the research in this thesis, and their personal concern for the welfare and development of their students has been continually demonstrated by the time and effort they dedicate to teaching.

I would like to thank John Aiken, Shih-Yin Lin, Emily Alicea-Muñoz, and Danny Caballero for the challenges they both posed to me and helped me to overcome. They have all been professional, personal, and academic inspirations to me, as well as indispensable companions at our conferences.

Anat Fintzi has been my partner and my guide for many years, including some rather difficult ones full of long-distance separation and uncertain reward. Her own academic work has informed mine, her statistical expertise has saved me from some grievous analytical errors, and without her loving support, it would have been impossible for me to write a single line of this thesis.

I would like to name the talented animators who helped to create our lecture videos, without whom our lecture series would not exist. They are Abhinav Uppal, Megan Llewellyn, Anthony Bandoly, Alana Lin, Olivia Ginn, Christina Chung, Nathan Olivares, Chris Syed, Zachary Wilken, and Evan Seitz. Special thanks goes to Evan for taking the lead on the Physics II video production team.

The developers and designers on the VIP team deserve credit for helping to build the SWAPR system for peer assessment. They are Stuart Freeman, Rachel Golding, Manley Roberts, Brad Reardon, Joseph Lesniak, Ayan Das, Christopher Wang, Titus Woo, Aadil Omar, Julian Barnes, and Austin Duncan. Without their help and that of the staff and faculty on the Georgia Tech VIP program, our peer assessment system would have likely remained unknown outside of our own small lab.

Finally, I would like to acknowledge all the colleagues, comrades, and creatures, known and unknown, on whose work I relied and from whose intellectual heritage I drew. My own efforts, as I see them, are just a single contribution to the vast social enterprise we undertake together.

Our work presented in this thesis was funded in part by the Bill and Melinda Gates Foundation, The Governor's Office of Student Achievement (Georgia), and the National Science Foundation (DUE-0942076). Our research methods were approved by the Georgia Tech Institutional Review Board (protocols H09382, H13124, H13302, and H13308).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

We designed and implemented a set of introductory mechanics laboratory exercises featuring real-world data-gathering, computational modeling, and video lab reports with peer assessment. Our goal in developing a peer assessment system was to create a valid substitute for instructor grading which could operate at scale, and to achieve learning goals related to student scientific communication and critique. We found our peer assessment system to be an adequate replacement for instructor grading of these lab report videos, and discovered that students learned to produce more expert-like assessments as the semester progressed, as demonstrated by a substantial rise in student-expert rating agreement.

Further investigation showed that this improvement in accuracy was related, at least in part, to the completeness (but not necessarily the correctness) of students' explanations of physics phenomena in their lab report videos. At the beginning of the semester, this completeness had no effect on the ratings which students would give each others' videos; at the end of the semester, explanation completeness was strongly correlated with lower student ratings. This correspondence between salutary communicative practices and peer assessment behavior indicates that our introduction of peer assessment was indeed effective at achieving communication-oriented learning goals.

Students' retrospective accounts of their experiences with peer assessment also produced some themes and trends that seemed to be corroborated in our quantitative research. All this work together, both quantitative and qualitative, will serve as a basis for continued research into modeling student engagement with peer assessment in the introductory physics classroom.

CHAPTER 1

INTRODUCTION

Physics instructors aiming to improve their students' learning outcomes will often turn to results from physics education research (PER). Conceptual surveys are commonly used in PER studies to judge the large-scale effectiveness of instructional reforms, [1, 2, 3] and a great deal of work is done with students to explore their conceptual understanding and to characterize the way they learn to solve physics problems [4]. New curricula and pedagogical practices are regularly assessed by examining their effects on student conceptual development [5, 6, 7]. All this, together, yields a rich source of practical and theoretical knowledge for physics instructors to improve learning.

The motivation for instructors to focus on conceptual development and problem-solving is clear. Each year, hundreds of thousands of undergraduate students take an introductory physics course [8], and 14,000 students graduate college with a science, technology, engineering and/or mathematics (STEM) degree [9]. Physics is a common prerequisite for fields as diverse as medicine, mechanical engineering, biology, and astronomy. Undergraduate students who take physics courses are expected to learn a huge number of physics concepts, to develop problem-solving strategies, and to pass written exams, all for the purpose of preparing them for the rigors of academia, for making productive contributions to industry, and for practicing informed citizenship.

Academia, industry, and citizenship, however, require other things beyond conceptual mastery and problem-solving expertise. Researchers in PER (and other fields of discipline-based education research) have demonstrated that developing expert-like learning attitudes, building disciplinary self-efficacy, and learning a host of interpersonal "soft skills" [10] are all important to academic outcomes, career achievement, and informed governance. When students are not given explicit instruction in these skills and attitudes but are nevertheless expected to demonstrate them in the physics classroom, researchers say these things form part of the "hidden curriculum" of physics [11]. Aside from leaving students unprepared for important parts of the actual practice of physics, well-intentioned instructors who unwittingly relegate important skills to the hidden curriculum can further mystify and demoralize students who may already struggle with physics concepts.

Frequently, the hidden curriculum of undergraduate physics includes the set of skills and attitudes related to formal physics communication and critique. Students giving presentations in higher-level STEM courses or research conferences and those working at internships in professional firms are expected to have developed the skills and attitudes necessary to compose, communicate, and critique scientific arguments in an expert-like way, yet are not often given relevant and explicit instruction in in their introductory courses. Introductory physics courses—which reach many students, and set the groundwork for their future experiences in physics—therefore pose a mostly unrealized opportunity to explore and develop students' physics communication and critique practices.

To help realize this opportunity, our research group developed and implemented a series of inquiry-based [12] laboratory exercises featuring out-of-classroom observational data collection, computational modeling, formal lab report video production, and calibrated peer assessment, along with a comprehensive set of introductory mechanics lecture videos and electronic homework assignments and quizzes, all capable of being administered entirely online. This was developed on top of our group's previous efforts implementing the *Matter & Interactions* [13] (M&I) curriculum in a large-enrollment course and extending it with enhanced computational exercises [14]. Aside from our instructional goals, the lab exercises (and their administration in various classroom contexts) were also intended to serve as a research tool for probing the development of scientific communication and critique among introductory mechanics students. In fulfilling the requirements of the exercises, physics students at Georgia Tech produced a rich dataset of laboratory videos, quantitative and qualitative peer assessments, and electronically-captured video engagement behavior (clickstream data).

1.1 Research Questions

This thesis aims to characterize our laboratory exercises and peer assessment system, and to analyze the way in which these exercises effected changes over time in students' physics communication and critique skills. The research questions pertinent to this thesis have to do with the way in which our students assessed their peers' lab report videos, the way in which students produced their own lab report videos, and how these practices changed over time, all with an eye to the development of students' scientific communication practices and critique skills. We look to students' assessment of their peers' video projects to characterize their changes over time in peer assessment behavior, and we assess the content of their video projects at the beginning and end of the semester to characterize their development of communication skills. We also expand and inform both investigations with student interviews and think-aloud [4] exercises to produce a working longitudinal model of student engagement with peer assessment. No widely-disseminated survey instrument specializing in student scientific communication or critique exists for introductory physics students, so the main source of quantitative data used in this thesis is from the lab report videos our students produced and the assessments they gave to each other's videos.

This thesis is organized as follows:

- Chapter 2 describes existing literature on classroom peer assessment and where our research fits in. This chapter describes the exact form of our peer assessment system, our motivations for designing it as we did, the classroom environments in which we implemented it, and the statistical methods we use in our later analyses.
- Chapter 3 addresses the question of whether our students' assessments of lab report videos are accurate enough to justify the use of our peer assessment system for assigning grades. We compare student assessments to expert assessments and quantify

how well they agree compared to other peer assessment systems described in the literature. We also describe a calibration algorithm we designed to increase the accuracy of our peer assessment system.

- Chapter 4 addresses the research question, *In what ways do students change in their assessment practices as they gain experience with assessing each other's lab report videos?* We describe a detailed comparative statistical analysis of student assessments and expert assessments of a set of lab report videos, and identify trends in student assessment practices over time.

- Chapter 5 addresses the research questions, *Why do student assessment practices change in the way they do? What changes in student attitudes and understanding effect these behavioral changes?* We examine a set of student clinical interviews and think-aloud exercises, and identify themes relevant to rating behavior expressed by students in their retrospective accounts of their experience with peer assessment.

- Chapter 6 addresses a two-part research question, *How do students' explanations of physics concepts in their lab report videos change over time, and how do these explanations influence the ratings which students give to their peers' lab report vidoes?* We describe a coding survey of the various kinds of explanations of physics principles offered by students in their lab report videos, and we identify through quantitative analysis a change in rating behavior associated with these codes.

- Chapter 7 summarizes the findings of previous chapters, offers a potential mechanism underlying the trends described in Chapters 4-6, describes our existing body of data, and outlines a future research agenda for developing a fuller model of student engagement with peer assessment.

CHAPTER 2

BACKGROUND AND METHODS

2.1 Introduction

In this chapter, we set the background for our implementation of peer assessment and our study of changes in students' peer assessment behavior over time. We summarize existing literature regarding novice/expert comparisons generally and peer assessment in particular, and identify the results relevant to our own design (particularly, the logistical and pedagogical implications of peer assessment). We also describe the classroom settings in which our peer assessment system was implemented. Finally, we describe our peer assessment system in detail, along with the statistical methods we use in later chapters.

2.2 Novice and Expert Comparisons

Our research into peer assessment categorizes students and instructors as novices and experts, respectively, and examines the behavioral and affective changes exhibited by novices as they receive instruction, engage in peer assessment, and (we hope) develop expertise in physics and in scientific communication and critique. Some discussion of existing work comparing novices to experts is therefore warranted.

Many psychologists and education researchers have investigated the behavioral and cognitive differences between novices and experts performing tasks in a wide variety of domains. A seminal finding in the field of novice/expert differences is described by Chase & Simon (1973), who studied the way in which novice and expert chess-players perceive chess configurations (*i.e.,* the configurations of chess pieces on a chessboard) and plan out future moves. [15] The researchers found that experts could recall valid chess configurations much more accurately than novices could after being shown the chessboard for five seconds; yet experts were no better than novices in recalling the configurations when pieces were placed randomly. The researchers attributed this effect in part to the role of "chunking" in recall and working memory; expert chess-players, they concluded, had developed a facility at extracting information from brief inspections of chess configurations by attending to the configurations of subgroups of pieces rather than by attending to the positions of individual pieces. These subgroup configurations constituted "chunks" of perceptual information that encoded the configuration of many pieces as a single element in working memory (and could be recalled as a single element from long-term memory). Novices, lacking an extensive long-term mental library of chess-configuration chunks, instead attended to and encoded the positions of individual pieces and therefore exhausted the capacity of their working memory before encoding the entire configuration. [15]

Later work found this theme of experts and novices attending to different features of a situation in the domain of physics. Chi *et al.* undertook a study of physics experts and novices categorizing physics textbook problems by type, and investigated how the two groups created categories and by what criteria they sorted problems into those categories. [16] The researchers discovered that physics experts tended to sort problems into categories based on fundamental physics laws, or other abstract features relevant to problem-solving strategies such as whether a physical system was described both before and after an interaction (suggesting *e.g.* that the solution might involve a conservation law). Novices, by contrast, would tend to create categories and sort problems into them based on "surface structures" and concrete terms literally present in the text of the problem, such as blocks, ramps, or the phrase "friction". [16]

Some researchers express the differences between novice and expert attention, chunking, and categorization in terms of "schemas", being "principle-oriented knowledge structures" encoded in long-term memory and activated by perceptual cues. [17] From this perspective, the most salient differences between novices and experts when faced with a problem to solve or an artifact to analyze are 1) the expert's abundance of deep, highly developed schemas in long-term memory and the novice's relative lack thereof, and 2) the expert's facility with identifying the perceptual cues relevant to activating appropriate schemas. Domain expertise, from the schema-oriented perspective, therefore consists in large part of the development and accumulation of schemas relevant to domain-specific tasks; and by corollary, effective instruction is that which helps learners to develop useful schemas. [18] Our own work in developing a rubric for peer assessments was, in a sense, an attempt to develop an instructional tool for assisting students to develop schemas for assessing peer scientific communication.

Research into schema development has yielded another concept potentially relevant for the work described in this thesis: expertise reversal. As described above, from the schema viewpoint, the task of successful physics instruction is to develop instructional tools and methods which help learners to develop mental schemas which are useful for performing various tasks in physics. These instructional tools may consist of scaffolded computer code, explanatory tooltips written on diagrams, auditory narration describing how one is supposed to approach a problem represented schematically on-screen, etc. However, researchers have found that certain schema-oriented instructional methods which improve learning outcomes for novices can *degrade* learning outcomes for experts faced with the same task. Sweller *et al.* [18] describe this "expertise reversal effect" in terms of working memory capacity and cognitive load. They conclude that including schema-oriented instructional methods such as those listed above necessarily imposes an additional cognitive burden on the learner. For novice learners, the degree to which the instructional method improves their own schema development easily outweighs the negative effects of having to attend to additional code comments, text labels, etc.—but for experts who already possess well-developed relevant schemas, the schema-development instructional methods are entirely redundant, and serve only to impose an additional cognitive burden. [18]

This research into expert/novice comparisons suggests that our qualitative exploration of peer assessment should consider the features of peer-produced artifacts to which novices and experts attend (and whether those features are "surface structures" or deeper abstract concepts), and that novices and experts may use our peer assessment rubric in different ways.

Having established some background into expert/novice comparisons generally, we now describe peer assessment in particular and explain the laboratory assignments we use to identify expertise and novicehood among our peer assessors.

2.3 The Motivations and Forms of Peer Assessment

Peer assessment is used as a teacher aid and instructional tool in classrooms around the world, in fields spanning law, science, humanities, medicine, and elementary education [19, 20]. While the actual frequency of the use of peer assessment in classrooms is unclear, a brief review of the literature indicates that peer assessment in its various forms has become a popular topic of study since at least the 1980s. The rise of online learning platforms have also come with an expansion of the role of peer assessment; two major massive online open course providers, edX [21] and Coursera [22], have both incorporated online peer assessment systems into their course offerings since they were founded in 2012. The ongoing adoption of different styles of peer *learning* in large classrooms has also brought with it a complementary expansion of peer *assessment*, as well [23].

Peer assessment in all fields involves students at roughly the same level of education evaluating each other's work. Beyond this basic similarity, though, peer assessment systems may differ widely, depending on the size of the course, the instructors' pedagogical goals, the student work being assessed, the age of the students, and many other factors. A meta-analysis by van Zundert *et al.* [20] examined the relationship between peer assessment features, student characteristics, and peer assessment outcomes.

Topping, 1998, [24] provides a comprehensive typology of peer assessment systems, categorizing them according to 17 variables. The huge variety of possible peer assessment systems affords instructors great flexibility in implementation, but also cautions us against drawing bold conclusions about peer assessment in general from a study of one peer assessment system in particular. Given this, researchers of must be careful to describe in detail the peer assessment systems investigated in their studies in order for their work to be situated in the literature, which we do later in this chapter.

While the different forms of peer assessment are numerous, the motivating benefits of peer assessment fall into a few broad categories [25]. Two of these benefits are relevant to our inquiry:

- Logistical—A peer assessment system which produces valid feedback (either formative or summative) allows open-ended exercises in large courses without prohibitive instructor or TA time commitments.

- Pedagogical—Students' own content knowledge can be deepened by exposure to other students' work, and peer assessment itself involves a set of critical skills and practices that can be developed with experience.

Chapter 3 covers the logistical implications of our peer assessment system, demonstrating the validity of the peer assessments in terms of peer/instructor agreement. Chapters 4 and 6 describe quantitative results pertinent to pedagogy—students learn to become more expert-like in their rating behavior as time goes on (a finding which corroborates the finding in Murthy, 2007, despite major dissimilarities in our studies), and this trend is associated with the completeness of the physics explanations their peers make. In Chapter 5, we undertake a qualitative exploration of these changes in student assessment behavior and describe the themes and trends we uncovered during student interviews, which informs our understanding of both the logistical and pedagogical qualities of our peer assessment system. Finally, in Chapters 6 and 7, we synthesize the findings of the previous chapters, proposing a mechanism underlying our observed quantitative trends in student rating behavior and setting an agenda for future research.

2.4 Previous Peer Assessment Research

Much quantitative research into peer assessment has been concerned with establishing the *validity* of the peer assessments, often through some measure comparing peer-produced scores to instructor scores. A widely-cited meta-analysis of peer assessment validity in higher education is provided by Falchikov & Goldfinch, 2000 [19]; we refer to this meta-analysis to compare our own results against existing findings when we examine the validity of our peer assessment system in Chapter 3.

Other researchers have gone beyond demonstrating the validity of peer assessment and have explored the relationships between peer assessment outcomes, methods, and population variables in some detail; van Zundert *et al.*, 2010, provide a meta-analysis of such studies and describe a set of variables relevant to peer assessment outcomes [20]. The authors, however, note that using the existing literature to draw inferences about cause and effect in peer assessment is difficult "because the literature usually describes [peer assessment] in a holistic fashion, that is, without specifying all the variables present in terms of conditions, methods and outcomes." [20] To avoid this deficiency, we describe our peer assessment methods and classroom conditions in the later sections of this chapter using a typology of peer assessment provided by Topping, 1998 [24], and we describe our peer assessment outcomes in great detail in Chapters 3-6.

Most of the studies included in both meta-analyses were concerned with establishing the overall validity of various peer assessment systems in different classroom settings and among different populations, and so were not designed to study *how* students learn to interact with a given peer assessment system through repeated engagement—and certainly not through repeated engagement in a physics classroom, specifically. This gap in the literature is not absolute; a preliminary study of peer assessment over time in a small-enrollment physics classroom (two sections of around 10 students each) is provided by Murthy, 2007 [26]. This study concerns students evaluating each others' homework using a rubric, and the author both demonstrates an improvement in student/instructor agreement over time and proposes a mechanism which may account for this improvement (*i.e.,* students' adoption of dimensional analysis as a solution-evaluation strategy). Our own work involves many more students ($N = 387$) in two different classroom settings (on-campus and online), and concerns peer assessment of rich video lab reports rather than written problem solutions.

2.5 Laboratory Design

Our introductory mechanics curriculum featured four laboratory exercises designed to be completed by individual students, each centered around a physical system exhibiting a particular type of motion.[1] The sequence of systems covered by our four labs was in line with the standard introductory mechanics canon; we began with zero-net-force motion in one dimension, then motion with a variable force in one dimension, then two-dimensional motion with a central force, and finally two-dimensional harmonic oscillation. The laboratory process described below is summarized visually in Figure 2.1.

2.5.1 Observational Data

In each lab, students either received or were instructed to gather a video of a system exhibiting a specified type of motion. Students were told to use the cameras on their smartphones or the webcams built into their laptops to gather this motion data—we provided no equipment to the students. This equipment-free design was a strict necessity for an online course where students never meet in person, but it was also a convenience for our on-campus sections, since we saved time and effort by not having to set up or maintain experimental apparatus.

For the first lab (zero net force), many students elected to take videos of car traffic on residential roads, or rolled balls or cans across their desks; for the second lab (1D variable force—drag from air resistance), students selected household objects of varying shapes

[1] A fifth laboratory exercise with no set topic and an unscaffolded peer review process is excluded from our analysis here.

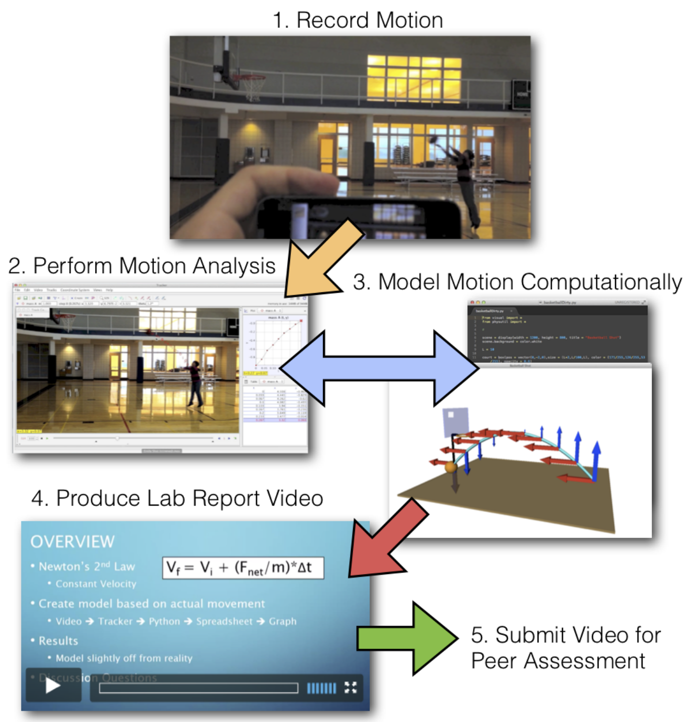


Figure 2.1: The laboratory-production process for students in our course. Students recorded video data of motion in their environment, derived a position vs. time plot for that motion using the Tracker motion analysis software, then modeled that motion computationally in VPython. Students would then each produce a roughly 5-minute-long lab report video in the style of a short symposium talk, where they would cover the fundamental physics principles used in the laboratory and compare their simulated motion to the observed motion. Each student would then submit their lab report video to their peers for assessment.

and densities and dropped them from a few meters' height. We provided students with trajectory data of a simulated astronomical system for the third lab (2D central force), and gave them a video of a mass swinging on a spring for the fourth lab (2D harmonic motion). In all labs, we either provided students with data or they collected it themselves with their own tools from their own everyday surroundings.

2.5.2 Motion Analysis

After collecting video motion data, students analyzed it in Tracker [27], a free Java-based video analysis tool available through the comPADRE Open Source Physics collection [28]. This analysis consisted of superimposing a coordinate system over the video and tracking the apparent position of the center of the moving object frame-by-frame, thereby producing time series data for the object's trajectory.[2] See Figure 2.2 for a visual example.

2.5.3 Computational Modeling

With this time series in hand, students then created a predictive computational model of the system in VPython [29] by modifying a provided minimally working program. Students computed the net force on the system and implemented Euler-Cromer integration of Newton's 2^{nd} Law; while the integration code was essentially unchanged from lab to lab, the computation of the net force in each lab involved different physical concepts such as drag, gravity, and Hooke's Law. Students then ran their computational models with parameters and initial conditions derived from their video analysis, and finally compared the model's output with their observed trajectories. Our intent in designing this part of the laboratory process was to provide students with experience creating, testing, and validating computational models of physical systems, practices which constitute the "third pillar [30]" of modern science (along with theory and experiment) but which are often absent from the introductory physics curriculum.

[2] Lab 3 (2D motion with a central force) deviated from this procedure by providing the students with trajectory data up front, cutting out the video analysis step.

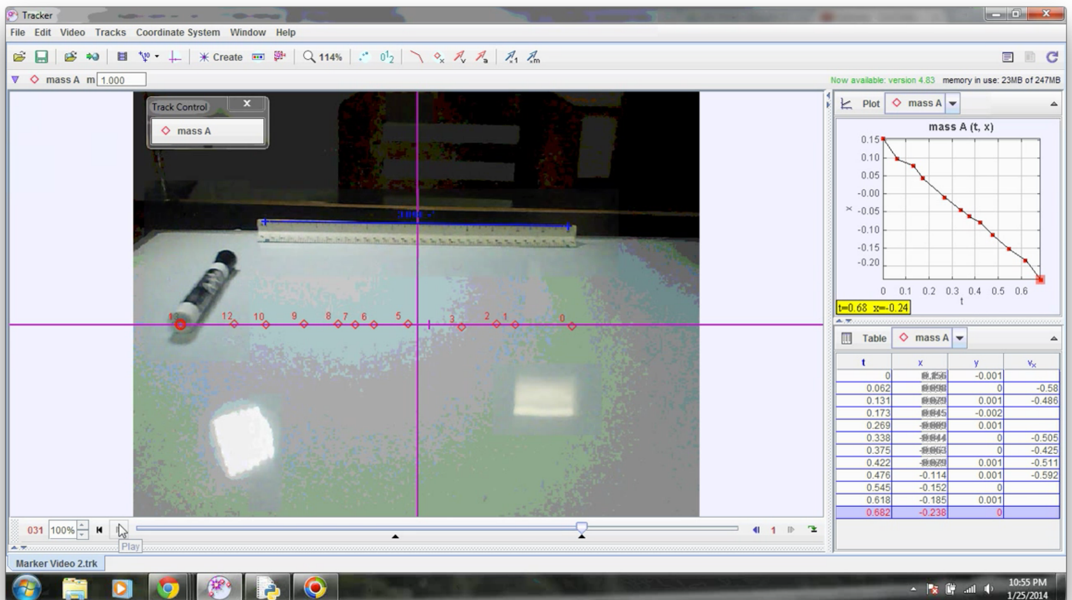


Figure 2.2: Example screenshot of motion tracking in Tracker, selected from a student lab report video. The tracked object—in this case, a dry erase marker—rolls across a table from left to right, with red points indicating the position of the marker at each frame. These points were set manually; note the error in the position vs. time plot in the upper-right corner. A ruler in the background serves as a reference object for the blue "calibration stick" that sets the scale of the video.

2.5.4 Video Lab Reports

After students completed their observations, motion analysis, and computational model, they each produced a five-minute lab report video in the style of a short contributed conference talk. Students were provided with a rubric for their lab report videos at the very beginning of the course, and were informed that they would use this same rubric to evaluate their peers' lab report videos throughout the year (and that their own videos would be evaluated in the same manner). We also provided students with several lab report videos from previous semesters to serve as examples. Students were free to choose the style and format of their lab report videos; the vast majority of them chose to make their videos in the "screencast" format with audio narration over static or animated slides.

Students were instructed to produce a lab report video length of five minutes' length, but to grant their peers an extra one-minute grace period during evaluation. Students were instructed to ignore anything after the six-minute mark in the videos they evaluated, and to deduct points accordingly. We had dual intent in setting this time limit; we wanted to set a reasonable time burden for students in both producing and evaluating their lab report videos, and we wanted to give them experience with the strict time limits characteristic of presentations at academic conferences. As with computational modeling, the formal presentation of scientific concepts within a given time limit is another practice expected of professional scientists but often absent from the physics classroom.

2.5.5 Rubric

Each peer or expert grade was calculated from a rubric consisting of five items rated on a five-point Poor-to-Excellent rating scale. Each possible rating was assigned a numerical score. During peer grading, our peer assessment system showed students the overall numerical grade corresponding to their selected set of ratings. Students were given an unlimited number of opportunities to revise their ratings before the assignment deadline.

An early version of our rubric excluded any assessment of the lab report video's production quality. The instructors believed that the rubric should reflect the goals of the course, and that developing good production quality (visual effects, audio cues, overall attractiveness, etc.) was a relatively unimportant goal compared to the development of students' physics content mastery and argumentation. However, it was later determined that production quality was still an important consideration in students' minds, and could potentially influence students' ratings of unrelated rubric items if they weren't given the means to specifically express their opinions about it. In the final version of the rubric (used for this study), we accommodated this consideration by including a production quality item on the rubric but assigning it relatively fewer points than the other four items.

See Appendix A for the full rubric used in this study.

2.6 Classroom Practices

2.6.1 Spring 2014—On-campus

In Spring 2014, we administered peer assessment in a "flipped"[31] on-campus classroom context at Georgia Tech The 338 students involved were enrolled in two experimental sections of a large-enrollment Physics I course (PHYS 2211). This course used the Matter & Interactions textbook [13].

For the two experimental flipped sections, we replaced live lectures with online video lectures we developed [32]. (See Figure 2.3.) Class time originally devoted to lecture was instead spent on group work and problem-solving.

Each student was required to conduct their own experiment and submit their own video lab report, but we did not deliberately discourage students from working together. During class time, students met in small groups ($\sim$4 students) to present and receive feedback on their in-progress lab exercises and lab report videos.

If students had a dispute with the peer grade they received for their lab report videos, they could avail themselves of a regrade by a teaching assistant (TA).

Figure 2.3: Sample still frames from our animated physics lecture videos. Clockwise from top left: 1. Visual metaphor for a physical system, its surroundings, and the distinction between internal and external interactions (by Scott Douglas) 2. The Milky Way Galaxy (by Megan Llewellyn) 3. Component representation of the weight force, with example (by Scott Douglas) 4. Visual classification of units of measurement and the physical quantities they measure (by Abhinav Uppal)

2.6.2 Summer 2016 & 2017—Online

In Summer 2016 and Summer 2017, we administered peer assessment entirely online in a Summer Online Undergraduate Program (SOUP) Physics I course. The SOUP course used the Matter & Interactions textbook and our online lecture videos, just like the on-campus flipped course. The entire course was conducted online; students met online periodically for group work and TA-led question/answer sessions. 21 students participated in the 2016 SOUP, and 28 in the 2017 SOUP.

Each student was required to conduct their own experiment and submit their own video lab report, but as in the on-campus course, students were not actively discouraged from working together. Once per lab, students met online through Google Hangouts [33] in small groups ($\sim$4 students, plus one TA) to present and receive feedback on their in-progress lab exercises and lab report videos.

As before, if students had a dispute with the peer grade they received for their lab report videos, they could avail themselves of a regrade by a TA.

2.7 Our Peer Assessment System

Having described our laboratory exercises and classroom settings, we now describe the actual peer assessment system we developed.

2.7.1 Type and Motivation

Here, we summarize the form of our peer assessment system using the Topping typology [24]. Peer assessment as practiced in our introductory mechanics course had the dual objectives of course scalability and student improvement in critique and communication of physics. Our peer assessment system *per se* focused on quantitative, summative assessment, with some qualitative assessment provided through peer comments and some formative assessment during small-group rehearsals. (In this context, qualitative feedback means written commentary, and quantitative feedback means numerical ratings; summative assessment is the high-stakes assignment of a grade, and formative assessment is low-stakes feedback meant to inform students' future work). Lab report videos were the objects of assessment. Peer assessment substituted for instructor assessment, and contributed to the assessee's final course grade. Assessment was one-way, unsigned (*i.e.,* the assessors were unknown to the assessee, but the assessee was known to the assessors), with no face-to-face contact during assessment. Assessors were in the same course as the assessee, which implies similar ability but not necessarily same year of study. Groups of assessors assessed individual assessees, but assessors all worked independently and were unknown to each other. Assessment took place outside of the classroom, was compulsory, and counted for the assessor's course credit.

Our decision to build our peer assessment system in this way was strongly informed by our classroom environments. Our peer assessment system not only had to work for our on-campus and SOUP courses described above, it also had to work for a massive open online course (MOOC); that is, it had to accommodate students who would be taking the course individually and entirely online from remote locations, where face-to-face meetings would be impossible and regular online peer group meetings for our MOOC students would be difficult to arrange due to differences in students' time zones and schedules.

We intended the laboratory exercises, the objects of peer assessment, to be a major component of the overall course grade to emphasize the importance and real-world applicability of those exercises, implying that the peer assessment would be summative and high-stakes. We also expected a substantial number of drop-outs and incomplete assessments from students in our MOOC course, which suggested a need for multiple assignees per lab report video to ensure that each got at least one peer assessor. These motivations and requirements set the scope of our particular form of peer assessment, which is described in the next section.

2.7.2 Our Peer Assessment Process

At the end of each lab, after students submitted their lab report videos, they began a peer assessment process which comprised a "training" phase followed by an "evaluation" phase. All students were required to complete both phases of peer assessment themselves, but we made no attempt to ensure that students worked alone.

We designed the training phase to improve the accuracy of assessments, and to increase student confidence in the assessment process. As regards accuracy, Falchikov & Goldfinch found that peer assessment systems are more likely to yield greater peer/expert agreement when students feel "familiarity with, and ownership of" the criteria used in assessment rubrics [19]. While we did not directly involve students in the design of our rubric, we did give them extensive practice with training-phase video assessments to increase their familiarity with the rubric criteria in particular and the peer assessment process in general.

This latter form of familiarity was, we believed, important for improving student confidence in the system. A motivating factor in the development of our peer assessment system was our conviction that the peer assessment system would receive worthwhile input from peers only to the extent that peers trusted they would receive fair grades from it. In other words, we believed that if peers did not think the peer assessment system would produce trustworthy grades, then the peers would not devote the necessary time and effort to their assessments to produce trustworthy grades (a self-fulfilling prophecy). To head off this potential pitfall, we developed an extensive training phase, explicitly emphasized the importance of the training phase to students in our course materials and classroom meetings, and provided all students with the guarantee they could always get a re-grade from an individual teaching assistant if they ever disputed the grade they received from peer evaluation. We considered the matter of training and student confidence so important that, in the final peer assessment design, we had more training videos per lab (4) than peer evaluation videos (3).

The training phase comprised four lab report videos from previous semesters, selected and assessed by instructors. During training, each student viewed each video online, assessed it, and submitted the assessment electronically, after which they were presented with the detailed instructor assessment for that video. We directed students to compare their own assessments with the instructors' before proceeding to the next training video.

The first two training videos in each lab were for practice only, and carried no grade. The latter two training videos—the "calibration" videos—carried a grade incentive; students received a small grade penalty for assessments with ratings far from the ratings given by instructors. These calibration videos were also used to compute a confidence rating for each student used to adjust their final peer grades in a process described in Chapter 3. A fifth, final calibration video disguised as a peer video was presented to students during the evaluation phase to further ensure the integrity of the assessment system (but, not being advertised as a training video, this fifth video likely did not have any effect on student confidence in the peer assessment system).

During the evaluation phase, students were assigned three random peers' lab report videos (and one disguised calibration video), and asked to evaluate them. Students could assess all the evaluation-phase videos at their leisure, in whatever order they chose. The peer assessment process is illustrated in Figure 2.4.

2.8 Quantitative & Qualitative Research Methods

Our research into our peer assessment system, being fundamentally a study of human subjects in naturalistic environments, makes use of analytical methods not normally encountered in other fields of physics. As such, some background discussion of the statistical tests and concepts used in this thesis is merited.

PER concerns itself with the study of the actual practice of physics instruction, a complex human enterprise amenable to a number of very different research methods. These research methods are traditionally sorted into *qualitative* and *quantitative* categories. We broadly define qualitative methods in PER as those which aim to produce an understanding

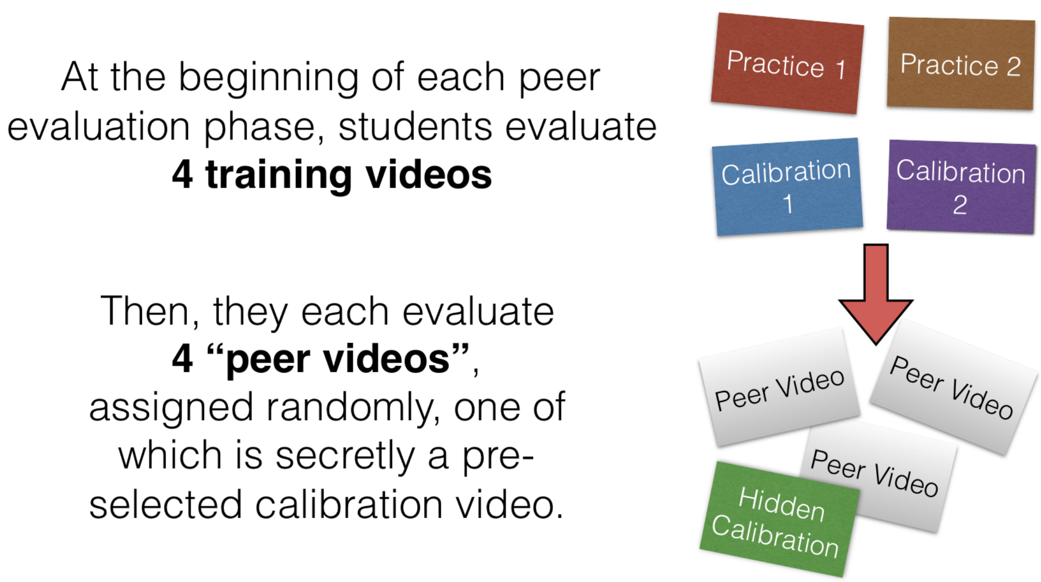


Figure 2.4: Illustration of the training (top) and evaluation (bottom) phases of peer assessment as implemented in our course.

of the subjective experience of physics education using data from not-strictly-controlled environments, while quantitative methods are those which aim to test an *a priori* hypothesis about a whole population in a way which controls for external variables [4]. These categories are not mutually exclusive, nor do they have sharply delineated borders; certain features common in qualitative methods (like small N values and rich verbal datasets) can also be found among quantitative methods, and the rigorous statistical tests typical of quantitative methods can also be used to achieve the exploratory, inferential research goals usually associated with qualitative research.

This thesis uses a mix of quantitative and qualitative methods to address the research questions listed in Chapter 1. The interview protocol which we use to conduct a qualitative analysis of student experiences with peer assessment is described in Chapter 5. Chapters 3, 4 and 6, containing mostly quantitative analysis, make extensive use of the statistical tests described later in this section.

2.8.1 Statistical Tests for Quantitative Analysis

A brief discussion of level of measurement, statistical significance testing, and effect size is needed to motivate our choice of statistical tests for our quantitative analyses.

The selection of statistical tests to be used in a quantitative analysis requires the identification of the *level of measurement* of the data under study, a typology which defines what mathematical properties are attributable to which data, and therefore which statistical tests are valid to use on those data. Formally, the level of measurement identifies a numerical structure to which a given empirical structure (*i.e.,* a "model of the data") is isomorphic [34]. As the word "level" implies, there is a natural ordering to this typology, with "higher"-level measurements possessing all the properties of all the levels "below" them (e.g., a particular statistical test may be said to be valid for measurements of "at least" a certain level).

From lowest to highest, the four canonical levels of measurement (as formulated by S.

S. Stevens [35]) are *nominal*, *ordinal*, *interval*, and *ratio*. *Nominal* measurements (sometimes called *categorical* measurements) produce values which differ from each other only by *name* (or *category*), such as species or nationality. The values produced by a nominal measurement have no natural order; like values can be grouped together with like values, but two unlike values defy any comparison other than "they are different". *Ordinal* measurements produce values which fall into a natural *order*, but with unknown or undefinable intervals between them. *Interval* measurements produce numerical values which fall into a natural order with well-defined intervals between them. The value "zero", however, has no natural meaning at the interval level and is placed arbitrarily, so ratios between interval measurements depend on the choice of zero-point and are therefore meaningless. Finally, *ratio* measurements produce values which fall into a natural order with well-defined intervals between them and a natural zero point. Because the zero point is natural, ratios of measurements are independent of the choice of units, and are therefore meaningful. The concept of level of measurement is mostly absent from formal physics instruction, where experiments seldom measure data of less than ratio level.

The data we examine in the quantitative research covered in Chapters 3, 4, and 6 is all of ordinal level or higher. In particular, student and expert *grades* and *scores* are interval measurements, so we can meaningfully compute and compare the means of different sets of grades and scores using Cohen's d (described below). Student and expert *ratings* are ordinal, so we compare sets with tests valid for that level of measurment: the Kolmogorov-Smirnov two-sample test and the Mann-Whitney U-test (also described below), neither of which require a set of values to have a definite mean.

Further description of these tests requires some background in statistical significance testing generally, the distinction between parametric and nonparametric tests, and the distinction between statistical significance and effect size.

2.8.2 Significance Testing

Every statistical test designed to discern differences between two sets of data poses a particular *null hypothesis* stating that sets A and B are not different in one way or another, such as “A and B are both drawn from a normal distribution with a single, well-defined mean” or “a value picked at random from set A is equally likely to be greater than or less than a value picked at random from set B”. Each test defines a procedure for calculating a *test statistic* relevant to that null hypothesis, such as a value proportional to the difference between the means of A and B, or the proportion of all possible pairs selected from A and B in which the A-value is greater than the B-value. Knowing the value of a test statistic alone is not enough to judge the veracity of the null hypothesis, though; having calculated the test statistic for a given A and B, one must then calculate a *p-value* representing the probability that that particular test statistic could have been arrived at by chance. This p-value is knowable because, given certain assumptions about A and B, the test statistic for a particular test is known to obey a particular probability distribution, and the p-value is just the *area under the curve of the test statistic probability distribution which is more extreme than the measured test statistic value*. For a given test, the p-values corresponding to a given N and a given test statistic are either tabulated or computable by formula or simulation. When the p-value is sufficiently low (traditionally $p < 0.05$), the test statistic is reported to be *significant* and the null hypothesis is rejected with a correspondingly high degree of confidence—in statistical significance testing, the rejection of the null hypothesis of a test is construed to be evidence of a real difference between A and B, not merely a difference due to sampling fluctuations. A full treatment of significance testing is beyond the scope of this thesis (for example, this description elides the difference between one- and two-tailed p-values); any introductory inferential statistics textbook will suffice for this purpose, such as Casella & Berger, 2002 [36].

2.8.3 Parametric and Nonparametric Tests

Statistical tests designed to compare sets of values fall into two general categories: "parametric" and "nonparametric" tests. The basic aim of any comparative statistical test, as described in the preceding section, is to determine if the two sets are dissimilar; parametric and nonparametric tests differ in the assumptions they make about the sets of values under consideration. Parametric tests assume that the sets under consideration are drawn from a population well-described by a normal distribution, and make claims about the dissimilarity of sets of values based on the dissimilarity of their means, variances, etc.—*i.e.,* the parameters which define a normal distribution. Non-parametric tests do not assume any particular distribution of the population underlying the data sets, and make claims about the dissimilarity of sets based on values which do not correspond to parameters of a normal distribution [37]. Note that this implies parametric tests cannot be meaningfully used with measurements of ordinal or lower level, since such measurements have no defined mean, variance, etc.; only nonparametric tests can be used for these data.

Beyond tests' applicability to different levels of measurement, there is also a trade-off between the *power* of a statistical test (its ability to discern a significant difference when N is small or the sets are very similar) and its *robustness* (its ability to produce correct results even when its assumptions are violated to some degree). As a rule of thumb, parametric tests make strict assumptions about data to gain power at the expense of robustness; non-parametric tests make fewer and looser assumptions, and gain robustness at the expense of power.

2.8.4 Effect Size

The *significance* of a statistical test is a statement about the likelihood that the result could have been arrived at by mere chance (*i.e.,* sampling variability). High significance indicates that whatever result is found is highly likely to represent an actual feature of the system being measured, and not a meaningless statistical fluctuation. Significance is related to sample size; a given measured difference between *e.g.* the means of two samples is more likely to be significant if those samples are large (high N) than if they are small (low N). This can lead to some confusion in reporting statistics, especially to general audiences, as *a highly significant difference* may seem to but does not necessarily imply *a big and important difference* [38].

The effect size of a statistical test, by contrast, is a statement about the *magnitude* of the effect which the test measures, and is independent of sample size. Many different tests of effect size exist, and the "magnitude" of the effect can be formulated in many different ways; Cohen's d is a common measure of effect size when comparing the means of two samples [39], while Pearson's correlation coefficient (often "Pearson's r" or just "the r value") is an effect size representing the degree to which two samples are linearly related [40]. Some significance tests use a test statistic that is itself an effect size, or is easily transformable into one.

The general procedure for performing statistical comparisons is to first select and perform a significance test and then, if the test produces a significant result, to determine the magnitude of the result by calculating the effect size. Effect sizes of insignificant results are usually not reported; we do not do so here.

2.8.5 Tests Used in This Thesis

Having covered this background, we now descripe the statistical tests we use in this thesis.

The Kolmogorov-Smirnov Two-Sample Test

A key analysis in this thesis (presented in Chapter 4) uses student ratings of peer video lab reports (an ordinal "Poor" to "Excellent" measurement) and the grades associated with those ratings (a measurement which is at least interval). This analysis makes extensive use of the Kolmogorov-Smirnov (KS) two-sample test, a very useful nonparametric statistical test for comparing two sets of ordinal-or-higher values [41]. (Not to be confused with the

Kolmogorov-Smirnov goodness-of-fit test).

The null hypothesis of the KS test is that two samples are drawn from the same underlying population distribution, but the KS test—like other nonparametric tests–does not make assumptions about the shape of that underlying distribution. The KS test statistic is the extremum of the difference between the two samples' empirical cumulative distribution functions, and varies continuously in magnitude from 0 to 1. This test statistic is sensitive to both the shape and the location of both sample distributions but is insensitive to scale. For these reasons, the KS test is suitable for ordinal data, and it also serves as a reasonable proxy for effect size.

The choice of the KS test in particular for our Chapter 4 analysis was motivated by its flexibility. We did not know in advance what sort of changes in student and expert ratings to expect (especially given that the means of the expert grades seemed to hold steady), and since the KS test is sensitive to differences between the shapes of whole distributions and not just *e.g.* their central tendencies, we believed the KS test would be useful for revealing unexpected changes in those distributions in this first, exploratory analysis.

However, the tabulated p-values for the KS 2-sample test are known to be conservative[3] for discrete distributions relative to continuous distributions [42, p. 435]. Given the relative discreteness of our ratings distributions, we were motivated to find a less conservative test for subsequent comparisons of central tendencies in Chapter 6 once we established in Chapter 4 that changes in central tendencies were indeed important using the KS test.

The Mann-Whitney U-Test

We compare distributions of ratings in Chapter 6 with the less-conservative two-sided Mann-Whitney U test (which measures the probability that an element selected from one group will be higher than an element selected from the other, and is therefore suited to com-

[3] In the context of inferential statistics, "conservative" tests tend to report a p-value that is larger than the exact, empirical value. More conservative tests are therefore less likely to produce Type 1 errors (false positives) but more likely to produce Type 2 errors (false negatives) when calculating significance.

paring the central tendencies of sets of ordinal data), [43] and used Pearson's Chi-Squared test on dichotomous variables (*i.e.,* binary variables, variables with only two possible values) to compare single-rating fractions across distributions [44, p. 34].

The null hypothesis of the two-sided Mann-Whitney U-Test when comparing two sets is that an element drawn at random from the first set is equally likely to be greater than or less than an element drawn at random from the second set. When this test indicated a significant effect in Chapter 6, effect sizes were calculated by dividing the U statistic by the product of the N's of the two groups; this value, known as the "common-language effect size", represents the likelihood that a random rating from the first set will be higher than a random rating from the second [43, 45]. The "first" and "second" sets that were fed into the SciPy Mann-Whitney U function [45] as (x,y) were arranged appropriately by examination; *e.g.,* if the post complete ratings appeared by examination to be typically higher than the post incomplete ratings and the two-sided U test indicated significance, then effect size would be calculated with the post complete ratings as x and the post incomplete ratings as y.

Pearson's Chi-Squared Test

In Chapter 6, we investigate whether the frequency of individual ratings on an ordinal poor-to-excellent scale such as "good" varies significantly over time or between students and instructors. This calculation implies a division of the data into responses which are good and those which are not good—a dichotomous variable. Pearson's Chi-Squared test is a significance test commonly used to compare unpaired samples of nominal/categorial data; we use the chi-squared test as implemented in SciPy [46].

Cohen's d

We use Cohen's d as a measure of effect size when comparing the means of peer and instructor grades in Chapters 3 and 4. Cohen's d is an effect size frequently computed when comparing the means of two samples (note that this means Cohen's d is only defined for data which is at least interval). The most common form of Cohen's d, and the one used in this thesis, is the difference in sample means divided by the pooled standard deviation of the two samples,

$$d_{ij} = \frac{\mu_j - \mu_i}{\sigma_{ij}},$$

where d_{ij} is Cohen's d for samples i and j, μ_i is the mean of sample i, μ_j is the mean of sample j, and σ_{ij} is the pooled standard deviation of the two samples—*i.e.,* the standard deviation of the set containing all measurements in sample i and in sample j together. [39].

CHAPTER 3

VALIDITY OF OUR PEER ASSESSMENT SYSTEM

3.1 Introduction

While our overarching research goals were more focused on the changes in peer behavior over time resulting from experience with our peer assessment system than on the validity of the system itself, we nevertheless needed to establish the validity of our system in the first place. We determined to do this by situating the results of our peer assessment system in the existing literature. In this chapter, we compare our results to the findings of Falchikov & Goldfinch's [19] definitive meta-analysis of 48 peer evaluation studies, in which two different measures of validity are reported: effect size, and correlation. These two measures are useful for meta-analyses comparing the results of different studies because they are commonly reported explicitly in those studies, and if they are not, they can be calculated from statistics that almost always are reported, like means and standard deviations.

By comparing our own effect size and correlation to those reported by Falchikov & Goldfinch, we demonstrate that our own peer assessment system produced results which were within the range of validity demonstrated by other systems, though somewhat on the low end. We explore peer assessment validity as it is described in the existing literature, discuss some possible reasons behind our acceptable but lower-than-average results, and describe our efforts to boost the validity of our system through algorithmic calibration.

3.2 Peer Assessment and Validity

The logistical benefits of peer assessment are particularly attractive to physics instructors leading large-enrollment courses, both online and on-campus; because every additional student producing work is also another peer assessor, the capacity of any peer assessment system automatically scales up with the number of students in the course. In principle, this permits the use of content-rich student exercises with feedback in courses where the assessment of such exercises would otherwise constitute an untenable burden for instructors.

Peer assessment would confer no logistical benefit at all, however, if the assessments were so untrustworthy that instructors had to spend many hours correcting them anyway. The main logistical concern of any peer assessment system is therefore validity—does the system actually produce instructor-like assessments? Previous research has demonstrated that well-designed peer assessment systems are capable of providing an adequately valid replacement for instructor grading in many fields [19, 25, 47]. These studies usually report validity in terms of agreement between grades given by instructors and by students to a sample of exercises. In their 2000 meta-analysis of 48 such peer assessment studies, Falchikov & Goldfinch [19] report overall good agreement between peer and instructor grades.

The authors of that meta-analysis identified several features of peer assessment systems which contribute to validity (*i.e.,* peer/instructor agreement), and which we paraphrase here:

- Peer assessments which involve an overall global judgment based on well-defined criteria produce better agreement than assessments which require marking several individual categories.

- Assessments of familiar academic products like essays and proofs are more likely to result in high agreement than are assessments of professional practice like simulated clinical examinations.

- Good experimental design yields higher agreement—*e.g.,* instructors who properly reported student population characteristics and used sufficiently large populations also tended to report higher agreement.

- Assessment by large (20+) groups of peers produce worse agreement than assessments by smaller groups or by individuals.

- Student involvement with the creation of assessment rubrics tends to improve agreement [19].

When discussing our measurements of the validity of our peer assessment in later sections of this chapter, we will refer to these factors as potentially explanatory of our results.

Other instructors have augmented their peer assessment systems with algorithmic weighting of peer grades [48, 49, 50] and peer-matching procedures based on machine learning [51] to further improve accuracy. A 2010 meta-analysis by van Zundert *et al.*[20] identified several studies [52, 53, 54] where the training of peer assessors was shown to be associated with better peer assessment outcomes. One of our previous studies [55] also demonstrated the effectiveness of training in improving the accuracy of peer assessment.

3.3 Our Peer Assessment System Compared to Others

Here, we report the ranges of two different measures of validity exhibited by other peer assessment systems in Falchikov & Goldfinch's meta-analysis [19] (effect size and correlation), and compare them with those same measurements for our own system.

3.3.1 Effect Size

Falchikov & Goldfinch's meta-analysis reported a range of Cohen's d [38] of -0.75 to 1.25 for effect sizes comparing peer grades with expert grades for a matched set of exercises, with a weighted mean effect size of -0.02. Recall from Chapter 2 that Cohen's d measures the differences between the means of two sets, expressed as a fraction of their joint standard deviation. If the two sets have identical means and finite standard deviations, d is zero. The authors of this meta-analysis cite the close-to-zero mean d as evidence that peer grades agree well with expert grades, on average.

Our own student/instructor effect sizes at the beginning ("pre") and end ("post") of the semester were $d_{\text{pre}} = 1.59$ and $d_{\text{post}} = 1.02$, respectively.[1] These effect sizes provide additional support for our conclusion that peer/expert agreement increased over time, but they also place our study at the very high end of the effect size range found by the meta-analysis. High positive values of effect size mean that our peers gave substantially higher grades than did the experts, and though this effect diminished with time, it was still large at the end of the semester.

3.3.2 Correlation

The authors of the meta-analysis also compared correlations between peer grades and expert grades, finding a mean Pearson's correlation coefficient of $r = 0.69$ with a range of $r = 0.14$ to 0.99, concluding that the relatively high mean r further supported the conclusion that peer and expert grades agreed reasonably well [19]. Here, the authors take high correlation to indicate high agreement between peers and experts. Our own data shows a peer/expert correlation $r_{\text{pre}} = 0.41$ and an $r_{\text{post}} = 0.29$; for the purposes of situating our findings within existing literature, these relatively low r values indicate lower-than-average peer/expert agreement that is nevertheless well within the range of agreement that other systems exhibit.

This measured decline in correlation would appear to indicate a declining peer/expert agreement over time, in contradiction with our findings with Cohen's d (and, indeed, our later findings generally)[2]. However, certain features of this analysis and our dataset make us hesitant to adopt this interpretation.

Pearson's correlation coefficient r, though commonly reported when comparing matched

[1] Whereas this study reports effect sizes using the formula for Cohen's d presented in Section 2.8.5, Falchikov & Goldfinch use a different formula provided by Cooper [56]. For our data, the two different effect size calculations produce a difference of only < 0.01, so we use the more familiar form of Cohen's d. Note, also, that this use of effect size is different than our usage in Chapter 4, where we use effect size to compare pre and post *differences* between peer and expert grades. Here, we use effect size to compare a set of student *grades* to a matched set of expert *grades*. Perfect peer/expert agreement would yield $d = 0$ with this method.

[2] Falchikov & Goldfinch report only a single value of r for the studies in their meta-analysis, so we do not know whether those other forms of peer assessment exhibited a similar trend in pre-post agreement.

samples of values, does not seem to satisfy the common sense of "agreement" as it pertains to peer/expert agreement (and which is used in the rest of this thesis). Pearson's r, which varies between -1 and $+1$, represents the degree to which a linear relationship exists between two matched samples x and y (in our case, x being the set of expert grades, and y the matched set of peer grades). $r = +1$, perfect correlation, implies a system where each $[x, y]$ pair lies exactly along a line $y = mx + b$ (where $m > 0$)—by taking perfect correlation to constitute perfect agreement *per se*, any such linear relationship will suffice.

However, the common sense of "perfect peer/expert agreement" implies *the particular* linear relationship $y = x$, which is not necessarily implied by $r = +1$; this sense of agreement suggests we should be concerned with *how far, on average, the peer grades are from their matching expert grades* rather than the degree to which peer and expert grades exhibit a linear relationship. A standard measurement for the average difference between two matched sets of values is the root mean squared error (RMSE), used often as an optimization parameter in model-fitting [57]. The RMSE for a comparison of two matched sets as used here is

$$RMSE_{x,y} = \sqrt{\frac{\Sigma_{i=0}^{N}(x_i - y_i)^2}{N}},$$

where x and y are two matched sets of values of size N. RMSE is zero when the sets are in perfect agreement ($x = y$) and positive with no upper limit otherwise, with higher RMSE meaning less agreement.

Our data may therefore deviate from perfect agreement, defined as student grades y being the same as expert grades x, in at least two different ways; either by hewing closely to some linear relationship different from $y = x$ (which would admit a high correlation but high RMSE), or hewing to $y = x$ only loosely (which would admit a relatively low correlation but lower RMSE). Whether low RMSE or high correlation is more important in determining peer/expert agreement is a matter of judgment not addressed by Falchikov &

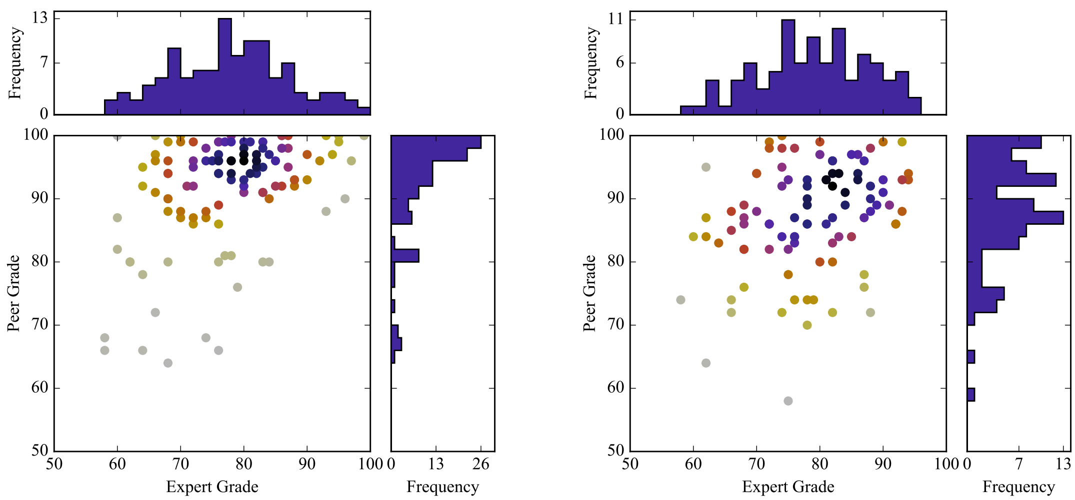
Goldfinch.

In our own data, we are in fact required to make this exact judgment. Our pre data (Figure 3.1a) exhibits an $r_{\text{pre}} = 0.41$ and an $RMSE_{\text{pre}} = 17.32$, while our post data (Figure 3.1b) exhibits an $r_{\text{post}} = 0.29$ and an $RMSE_{\text{post}} = 13.63$. That is, over the course of the semester, our peer/expert agreement decreased as measured by correlation, but increased as measured by RMSE. Since RMSE concords with the sense of "peer/expert agreement" we used in developing and executing our research agenda (where we take perfect agreement to mean $y = x$), and since other measurements of agreement discussed in later chapters also report an increase in agreement over time, we will disregard our measured decline in r as spurious, likely an artifact of the forced narrowing caused by the unusually high "pre" peer grades being constrained by the ceiling of the grade scale.

This still, however, leaves us with the matter of our relatively low agreement as measured by d and r *overall*, which we discuss in the following section.

3.3.3 Reasons for Relatively Low Agreement

Our comparatively low overall agreement as measured by d and r deserves examination. Our own peer assessment system featured two prominent factors identified by the Falchikov & Goldfinch as contributing to low peer/expert agreement, mentioned earlier in this chapter. Firstly, the authors conclude that peer assessment studies which have students assess "academic products and processes" of the sort they "have experienced for much of their formal education" tend to have higher agreement than studies where students assess "professional practice which requires them to learn a new set of skills." [19] The lab report videos our own students produced were novel products, intended to include some practical elements of professional scientific communication. Students had to learn video production techniques and new communication skills and were unlikely to have previously conducted a formal critique of such work in the classroom, so it seems fair to characterize our own peer assessment exercises as assessments of "professional practice" as defined by Falchikov &



(a) Pre

(b) Post

Figure 3.1: Matched peer/expert grades for 204 peer lab report videos selected from the beginning ("pre") and end ("post") of the semester. Scatterplots indicate the peer and expert grade for each video; histograms along each axis indicate the distributions of peer and expert grades separately. Color coding of scatterplot points indicates density of points. Pre peer/expert grades show a correlation $r_{\text{pre}} = 0.41$ and a root mean squared error $RMSE_{\text{pre}} = 17.32$, while post peer grades show $r_{\text{post}} = 0.29$ and $RMSE_{\text{post}} = 13.63$. Note the preponderance of scatterplot points above the main diagonal; this shows students tended to assign higher grades than experts to the same videos.

Goldfinch.

Secondly, the authors of the meta-analysis conclude that studies where students feel more "ownership" of the assessment criteria (*e.g.,* where students are directly involved in developing the grading rubric) tend to exhibit higher agreement [19]. Our own rubric development process did not involve the direct input of students.

Given the presence of both of these features, it's perhaps not surprising that our own system produced peer/expert agreement on the low end of the scale compared to other peer assessment systems (though not so low as to indicate serious problems with validity). In order to improve the accuracy/validity of peer assessments and to increase student confidence in the peer assessment system itself, we designed our peer assessment system to include calibration.

3.4 Calibration

The overall goal of calibration is to improve the accuracy of peer assessment (relative to expert assessment) by assigning a sort of confidence rating to each student in a training or calibration phase before the actual peer assessment. In a typical calibration scheme, students assess a set of documents that have been previously assessed by course instructors or other content-matter experts (the *calibration set*). The students' and experts' assessments of the calibration set are then compared, and each student is assigned a calibration score depending on the degree of agreement between their assessments and the experts'. Once each student is so calibrated, they all then assess each others' documents. Multiple students will typically assess each student's document; these assessments will then be combined to produce a final peer grade for that document. In calculating this final grade, more weight will be assigned to the assessments of students with higher calibration scores.

Calibration schemes vary in their specifics, such as the method by which calibration scores are calculated, and the method by which calibration scores are taken into account during the final grade calculation. These methods together constitute the *calibration al-*

gorithm. The popular commercial product Calibrated Peer Review™[50] is used in many different classrooms for the assessment of written essays; they do not make their calibration algorithm publicly available (Arlene A. Russell, personal communication, 8 February 2018).

3.4.1 SWAPR 1.0

In the interest of transparency and research utility, we developed our own open-source peer assessment system which we named "Statistically Weighted Aggregate Peer Review" (SWAPR). Version 1.0 of SWAPR consisted of a set of locally-run scripts written in Python [58] that implemented the calibration algorithm, and online assignments hosted on Webassign [59] and written in the Webassign-specific version of Perl [60] that served the lab videos to students, collected student assessments, and managed the gradebook. We used SWAPR 1.0 for all courses represented in this dissertation.

3.4.2 SWAPR 2.0

At the time of writing, Version 2.0 of SWAPR [61] is under development by this author and a team of students at Georgia Tech in the Vertically Integrated Projects (VIP) program [62]. SWAPR 2.0 is designed to be a full-featured free and open source calibrated peer assessment system, requiring neither integration with any existing platform nor the purchase of a site license. It features a backend written in node.js [63] with a PostgreSQL database [64] and a client written in AngularJS [65], and handles the submission, distribution, calibration, assessment, and reporting of students' work.

3.4.3 Algorithm Development

In designing our calibration algorithm, we sought to fulfill two goals:

- Transparency: Our algorithm should be repeatable and simple enough that students are able to fully understand and follow along with every step in the calibration process

- Accuracy: Our algorithm should produce final grades for students' videos that agree with expert grades more closely than alternative methods

Transparency

Again drawing on the finding by Falchikov & Goldfinch that peer assessment systems are more likely to yield good outcomes when students feel "familiarity with" parts of the system [19], we wanted our calibration algorithm to be strictly repeatable and use easily comprehensible calculations so as to avoid mystifying students with their calibration scores and final grades. This ruled out the commercial Calibrated Peer Review™system, which would have required the purchase of a site license to reveal its proprietary algorithm. This also ruled out algorithms involving sophisticated machine learning methods, Bayesian inference, or Monte Carlo optimization or other probabilistic processes, such as have been developed by other teams [48, 51].

Ultimately, as described in the following section, we arrived at a calibration algorithm that used only simple arithmetic and weighted sums. When we finalized our calibration algorithm for use in Spring 2014 and later offerings (as covered in the next section), we provided students with a document fully describing the calibration algorithm and providing a simple example of both the calibration score calculation and the final score calculation. (See Appendix B)

Accuracy, Development, and Optimization

In achieving accuracy, we aimed to develop a calibration algorithm that produced greater agreement with experts' final grades than the built-in peer assessment system on Coursera, which at the time of the development of SWAPR 1.0 was uncalibrated and produced a final grade by taking the median of the peer grades [66]. We began algorithm development using the set of peer assessments from the Fall 2013 on-campus sections, and tested the accuracy of several candidate algorithms by comparing their output with the actual expert scores for a set of videos.

The candidate algorithms relevant for this thesis were $median$, $mean$, $BIBI\_1$, and $offMean_1$:

- $median$ and $mean$ used no calibration scoring and calculated the final score by taking the simple median or mean of raw peer scores, respectively.

- $BIBI\_1$ worked by adding points to each student's calibration score for each time the student's ratings for a given video agreed with or were adjacent to the corresponding expert rating (*e.g.,* when the experts gave a fair, and the student gave a good), then calculated final peer scores with a weighted arithmetic mean of student scores, using the calibration scores for weights.

- $offMean\_1$ extended $BIBI\_1$ by also calculating the mean difference between the student's scores and the corresponding expert scores, then subtracting this difference from each student's score before calculating the final weighted arithmetic mean of their scores.

Variations on these algorithms were also considered, including algorithms which collapsed the 5-point rating scale into a 3-point scale, but initial testing showed that these variants did not produce substantially different results than the four algorithms described above.

The algorithm development dataset contained assessments from 387 students of 20 training videos, 12 of which comprised the calibration set and 8 of which comprised the practice set. Each student in our dataset had assessed each video. We used the calibration set to calculate calibration scores for each student using each candidate algorithm, then randomly generated 10,000 unique sets of 2 to 10 students[3] to simulate the effects of random

[3]For groups of size 2, there were not enough combinations of valid students to produce 10,000 unique groups, so 3741 unique groups of 2 were generated instead. Valid students were those who rated every single rubric item for all the practice and calibration videos for labs 1-4.

peer grouping and different group sizes. We then calculated the final set of rubric scores assigned to each video in the practice set by the students in each random group according to each candidate algorithm. For each candidate algorithm, we compared each peer group's final scores for each video, p_{aNgvr}, with the expert scores for that same video, e_{vr}, by taking the square root of the sum-of-squares error for that group,

$$err_{aNg} = [\Sigma_{v,r}(e_{vr} - p_{aNgvr})^2]^{\frac{1}{2}},$$

where a indicates the candidate algorithm, $N \in [2, ..., 10]$ indicates the number of students in each randomly-generated peer group, $g \in [0, ..., 9999]$ indicates the specific group, $v \in [0, ..., 7]$ indicates the practice-set video, and $r \in [0, ..., 4]$ indicates the rubric item.

The mean error for groups of size N for a particular algorithm a, then, is

$$err_{aN} = \frac{1}{len(g)}\Sigma_g[\Sigma_{v,r}(e_{vr} - p_{aNgvr})^2]^{\frac{1}{2}},$$

where $len(g)$ is the number of simulated groups of that size (in this case, 10,000).

Plotting these error values for each candidate algorithm as a function of group size yielded three results, as shown in Fig. 3.2. The first result was that, as one might expect, larger peer groups produced on average lower error for every algorithm except $median$. For $median$, even-numbered groups of students tended to produce lower error than odd-numbered groups; this is due to the simple reason that the median value of an even-numbered set is defined to be the mean of the central two values, so for even-numbered groups $median$ behaved more like $mean$, which was a better performer overall.

The second result was that the $BIBI\_1$ algorithm, which included a weighted mean calculation, did not substantially outperform $mean$, which consisted of a simple unweighted mean.

The third result, visible in Fig. 3.2, is that $offMean\_1$ algorithm was the best per-

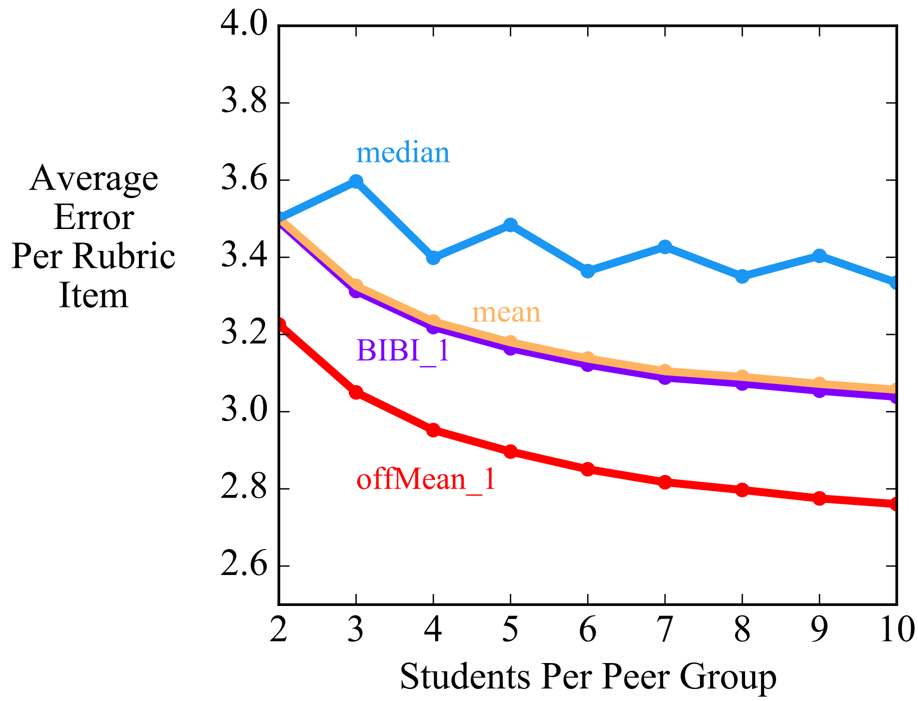


Figure 3.2: A simulation of different calibration algorithms using data from Fall 2013 peer assessments reveals that $offMean\_1$ produces the least error when comparing peer scores to expert scores. All tested algorithms generally tend toward lower error with larger peer groups, with $median$ exhibiting higher error for odd-numbered groups and lower error for even-numbered groups atop this general trend.

former of the candidate algorithms we tested, yielding about 0.3 points (out of 20) of improvement in accuracy for each of 5 rubric items. Assuming that this accuracy compounded across rubric items, this meant we could expect the $offMean\_1$ calibration algorithm to yield about 1.5 points of additional accuracy on each peer assessment over simply taking the median peer score, *à la* Coursera.

This was not a large improvement, but it was in a sense free; the students would be assessing the practice and calibration videos anyway as part of their training, the calculations necessary for finding calibration scores were computationally trivial, and one method or another had to be chosen to calculate final scores for the videos. In the end, we chose to use the $offMean\_1$ algorithm in the following term (Spring 2014) because it was the most accurate algorithm we had, and even a few points' worth of additional accuracy could mean a lot to students who were close to the thresholds between letter grades.

3.4.4 Algorithm Optimization

We ran SWAPR 1.0 using the $offMean\_1$ calibration algorithm for our Spring 2014 course. At the end of the course, we repeated the simulation protocol described above for $offMean\_1$ and $median$ with $N = 3$ to test whether our calibrated peer assessment system outperformed the uncalibrated Coursera alternative.

Using the calibration set to calculate calibration scores and the practice set to produce test results, and defining "agreement" to be when the calibrated final peer score and the expert score were within 2 points of each other, we found that our calibration algorithm raised peer/expert agreement from 33% to 45%, reduced the proportion of peer overscoring from 51% to 40%, and raised the proportion of peer underscoring from 0% to 4%. See Fig. 3.3. Once again, our calibration system produced a moderate improvement in accuracy.

3.5 Conclusion

Our peer assessment system demonstrated acceptable but lower-than-average validity (*i.e.,* peer/expert grade agreement) compared to other peer assessment systems. This comparatively low agreement may have been due to the unfamiliar and rather subjective nature of the lab report videos under review, and to the fact that students were not directly involved in the development of the rubric. Nevertheless, the overall validity of the grades produced by our peer assessment system ended up in the range demonstrated by other such systems.

In the following chapters, we investigate the quantitative changes in student-expert grade agreement over time in multiple levels of detail (Chapter 4), examine the themes and trends expressed in students' retrospective accounts of their experience with our system (Chapter 5), and tie these trends to students' video production choices (Chapters 6 and 7).

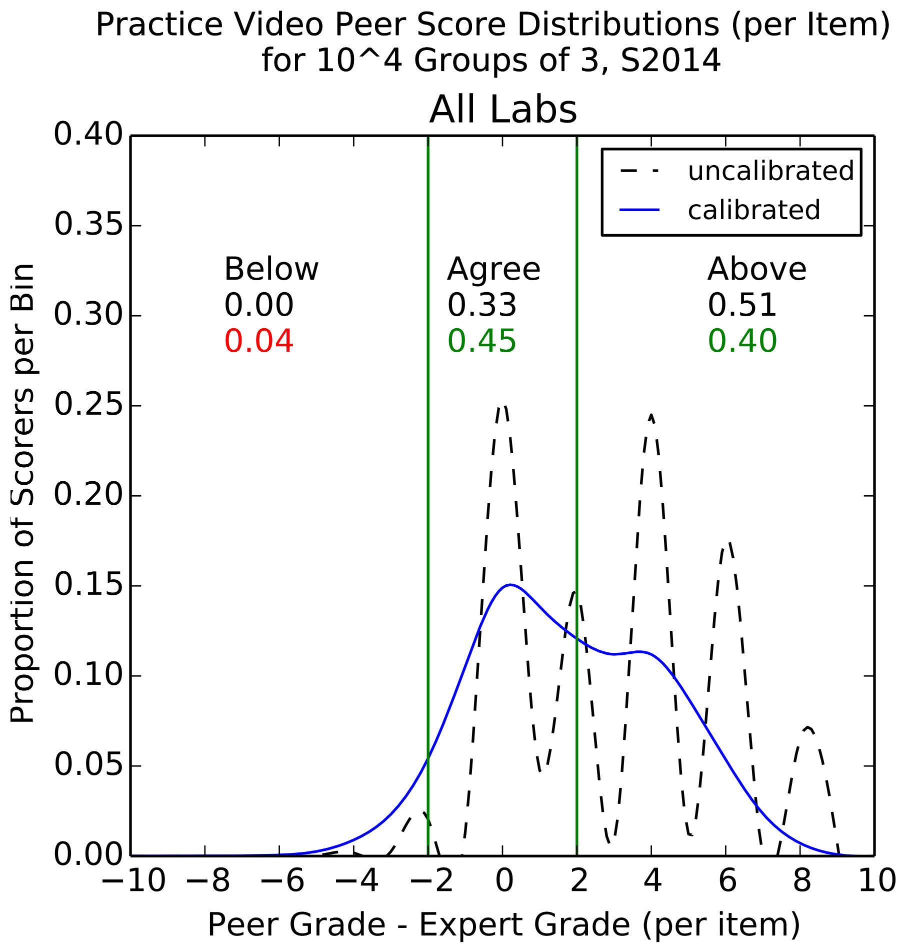


Figure 3.3: Comparing uncalibrated scores ($median$) with calibrated scores ($BIBI\_1$) for Spring 2014 practice videos shows an improvement in peer/expert agreement from 33% to 45% with calibration. Students tended to give higher scores than experts; with calibration, the proportion of peer scores that were higher than expert scores was reduced from 51% to 40%. Students seldom gave lower scores than experts; calibration raised the proportion of peer scores that were lower than expert scores from 0% to 4%. Overall, calibration produced a moderate improvement in peer/expert score agreement with no additional effort on the part of the students.

CHAPTER 4

STUDENT RATING CHANGES OVER TIME

This chapter presents our first look into the pedagogical aspects of our peer assessment system (see Chapter 1), and a shift away from the logistical analysis of Chapter 3. We quantitatively compare peer and expert assessments for a common set of student-produced lab report videos and describe their rating changes over time, and we compare student ratings of a set of videos common to several course offerings to establish the generalizability of our results across semesters.

A previous study of ours demonstrated an improvement in student/instructor agreement over time for a small sample of lab report videos [67]; the work in this thesis extends this original finding to a much larger set of videos and explores the underlying causes and mechanisms of that trend. This chapter presents findings we have published previously [68].

4.1 Introduction

In the course of performing their laboratory exercises and peer assessment as described in Chapter 2, our students produced an extensive set of lab report video assessments amenable to large-scale quantitative analysis. Our first foray into understanding the pedagogical aspects of student engagement with peer assessment was a quantitative analysis of a subset of these assessments as compared with matching expert assessments, which we report here.

We designed our quantitative investigation into lab video peer assessments as a pre/post study, with Lab 1 assessments constituting the "pre" dataset and Lab 4 the "post". We did this to set a reasonable limit on the scope of our investigation of peer assessments, and to align it with our other pre/post interventions like our administration of the FMCE [1] and student interviews.

In this chapter, we report separately our quantitative analysis of student assessments of the 10 pre and post training videos (which were assessed by all students) and of 204 selected pre and post peer videos (which were each assessed by a unique set of 3 students; see Section 2.7 for a full description of the peer assessment process). We find evidence of increased student-expert rating agreement as the semester progressed, both among the training videos and the peer videos.

4.2 Peer Ratings of Peer Videos

Here, we compare overall expert and peer grades of physics lab report videos selected from the beginning and end of the Spring 2014 on-campus session. All grade data in this chapter is gathered from this session unless otherwise noted. We omit in this section a comparison between average peer and expert grades of training videos, given the high variability of expert grades and small number of training videos; in the next section, we address these deficiencies by reporting *differences* between peer and expert grades of training videos (grading scheme a in Figure 4.2) as well as peer videos (grading scheme b).

We find that the difference between student and expert grades of peer videos diminishes over the course of the semester, and we follow this trend in this chapter through three increasingly fine-grained levels of quantitative analysis. As it so happens, the progression toward student/expert agreement we had previously detected among the training videos [67] was—broadly—also found to obtain among a large random sample of peer videos.

4.2.1 Peer and Expert Grades

Our highest-level examination of student and expert evaluation behavior—a simple comparison between overall grades—reveals that our students gave significantly lower grades to their peers' lab reports on average at the end of the semester ("post") than they did at the beginning ("pre"), while average expert grades for those same lab reports held steady. Mean peer pre and post grades (respectively, 90% & 86%, $p \ll 0.05$, Cohen's $d = -0.45$)

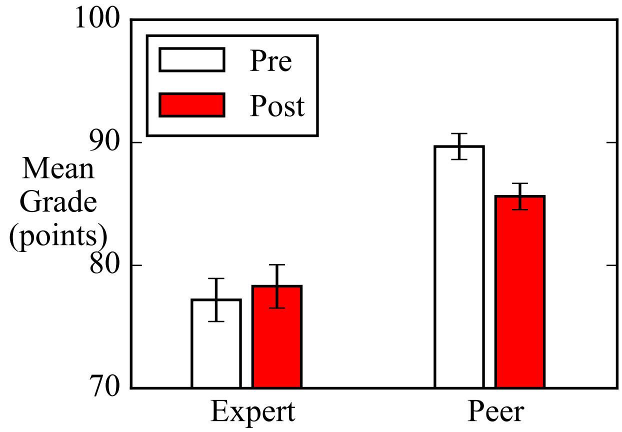


Figure 4.1: Peer assessments of physics lab report videos taken from the beginning (pre) and end (post) of the semester show that the mean peer grade becomes significantly lower and closer to the mean expert grade by 5 points, about half a letter grade on a standard 100-point grading scale ($p \ll 0.01$, Cohen's $d = -0.45$). A two-sample KS test shows distributions of expert grades do not significantly change ($p = 0.86$). Means are determined from 565 peer grades of a sample of 204 physics lab report videos selected from the first and last labs of the semester, along with expert grades for each of those videos. Error bars show 95% confidence intervals.

were around a full letter grade higher than mean expert pre and post grades (respectively, 77% & 78%, $p > 0.05$). Figure 4.1 shows the mean peer and expert grades for 204 lab reports submitted by peers at the beginning and end of the semester, each of which was assigned to three different peers for grading.

That the mean expert grade of students' lab report videos did not change significantly over the course of the semester should not be taken as evidence that the students did not improve in their physics understanding or in their laboratory skills. Unlike survey instruments, which are typically administered unchanged between pre and post, the labs were for-credit assignments situated in the natural progression of an introductory mechanics course. The last laboratory assignment of the semester involved a more complicated system than did the first (two-dimensional harmonic oscillation vs. one-dimensional constant-velocity motion, respectively). The last lab also required students to perform an analysis of mechanical energy, which the first lab did not. Since the physics content and methodological require-

Table 4.1: Differences between peer and expert grade for all peer grades of two groups of physics lab report videos (Group A, 204 randomly selected peer videos, 3 peers per video: Group B, 10 training-phase videos, 338 peers/video). Distributions of differences are compared with a KS 2-sample test, with effect sizes calculated by Cohen's d. Variances are compared with a Levene test. Standard deviations are reported instead of variances for clarity.

Group		Pre	Post	Significance	Effect Size
A	N_{pairs}	305	261		
	mean	12.38	7.30	$p \ll 0.01$	$d = -0.51$
	std	9.23	10.86	$p = 0.05$	—
B	N_{pairs}	1489	1163		
	mean	11.94	6.33	$p \ll 0.01$	$d = -0.53$
	std	10.57	10.57	$p = 0.29$	—

ments of the last lab were more difficult than those of the first, we view the stability of expert grades as evidence that students were overall able to stay on par with increasing instructor expectations.

The decline in mean peer grade indicates that peer evaluation behavior did change meaningfully over the course of the semester, and the relative stability of expert grades suggests that expert evaluations can serve as a useful baseline against which to compare peer grades.

4.2.2 Peer and Expert Grade Differences

Our next level of analysis involves matching peer grades to expert grades and taking the difference, rather than just comparing the means of the overall distributions of grades. We find that peer grading became more accurate over the course of the semester, but not more precise. In Figure 4.2, high peer-expert agreement would be indicated by narrow distributions centered on zero. Overall, the distributions of grade differences shifted significantly toward zero over the course of the semester, but they did not significantly narrow. See Table 4.1 for full tabulation of results.

We hypothesize that these results may indicate peers' experiences during the semester taught them that experts tended to give lower grades than peers did (accounting for the shift

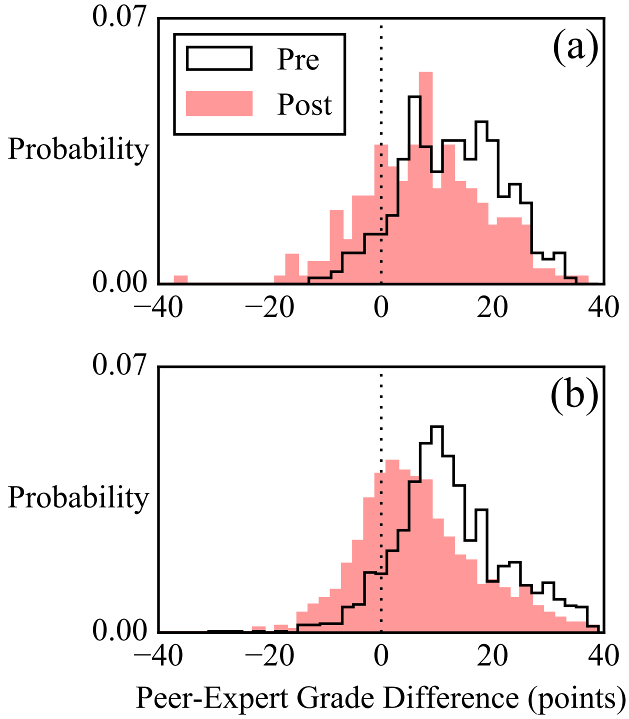


Figure 4.2: Distributions of differences between peer and expert grades show that peer grades become more accurate, but not more precise relative to expert grades. Two distinct grading schemes are considered: (a) In grading scheme one, many peer videos are each graded by few peers. (b) In grading scheme two, 10 training-phase videos are each graded by all peers. The mean grade differences shift significantly toward zero over the course of the semester, indicating peer grades increase in accuracy. The widths of the distributions do not change significantly, indicating peer grades do not become more precise. See Table 4.1 for N-values and full tabulation of results.

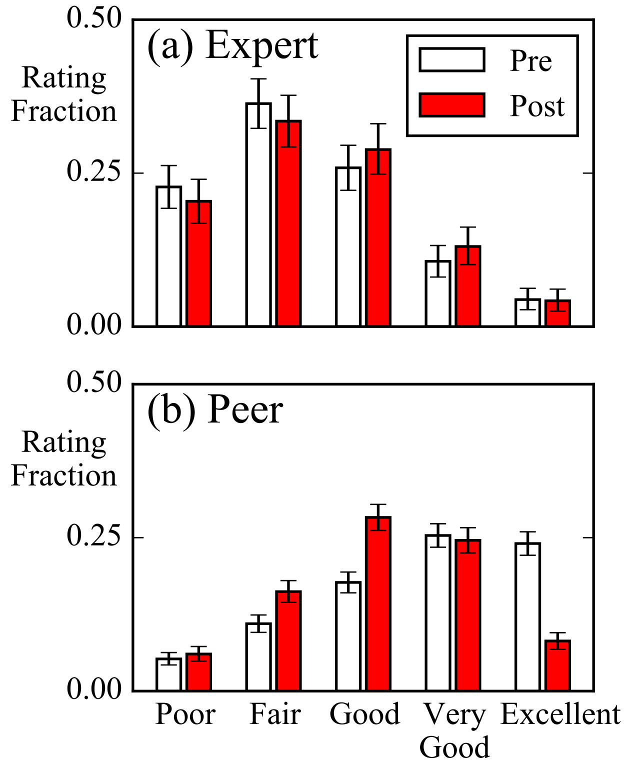


Figure 4.3: Peer rating distributions (a) change dramatically from pre to post, while expert rating distributions (b) do not. Five ratings on a five-point "Poor" to "Excellent" scale determine the numerical value of each peer or expert grade. "Excellent" changes from being among the most frequent peer ratings to among the least frequent, and "Good" ratings become substantially more frequent. Figure shows 2820 peer ratings of a sample of 204 physics lab report videos selected from the first and last labs of the semester, along with expert ratings for each of those videos. Error bars show 95% confidence intervals.

toward zero in Figure 4.2 and the lowering of the mean peer grade in Figure 4.1), but did not help them develop any further expert-like rating behavior (accounting for the lack of narrowing in Figure 4.2).

To inform this hypothesis, we examined matched peer-expert grades at a yet lower level by breaking them into their constituent "Poor" to "Excellent" ratings.

4.2.3 Peer and Expert Ratings

We compare the distributions of ratings underlying the grades given by peers and by experts at the beginning and end of the semester (Figure 4.3).

Ratings Overall

We find that the overall distribution of expert ratings is relatively stable across the semester. This means that expert ratings can, like expert grades, serve as a useful basis of comparison when describing student ratings. Experts give “Fair” ratings most frequently, and give comparatively few “Excellent” ratings, in contrast to peers who begin the semester giving more “Good” and “Excellent” ratings than any other rating.

By the end of the semester, the distribution of peer ratings had shifted substantially. The fraction of peer “Excellent” ratings greatly diminished, and the fraction of peer “Good” ratings greatly increased. No such trend was present among “Poor” and “Fair” ratings at the lower end of the rating scale; those rating fractions exhibited insignificant or relatively small changes, respectively. The “Very Good” rating fraction also did not change significantly.

Ratings for Physics-Content Rubric Items

At our final level of quantitative analysis, we compared the same sets of peer and expert ratings as above, but broken down by rubric item. The five items on the rubric comprised three physics-content and two non-physics-content items, encompassing the whole range of instructor expectations for lab video production. Items 2, 3, and 4, the physics-content items, asked the reviewer to assess—respectively—the author’s explanation of the physical model relevant to that lab, their discussion of their computationally simulated predictions versus their observations, and their overall grasp of basic physics concepts. For the full text of the rubric, see Appendix A.

Among the physics-content items, the peer ratings changed more than the expert ratings over the course of the semester, in line with our other findings regarding the relative stability of expert grades and ratings. In all three cases, the value of the two-sample KS test statistic (a proxy for effect size) for pre/post comparisons of student ratings are larger than for pre/post comparisons of expert ratings.

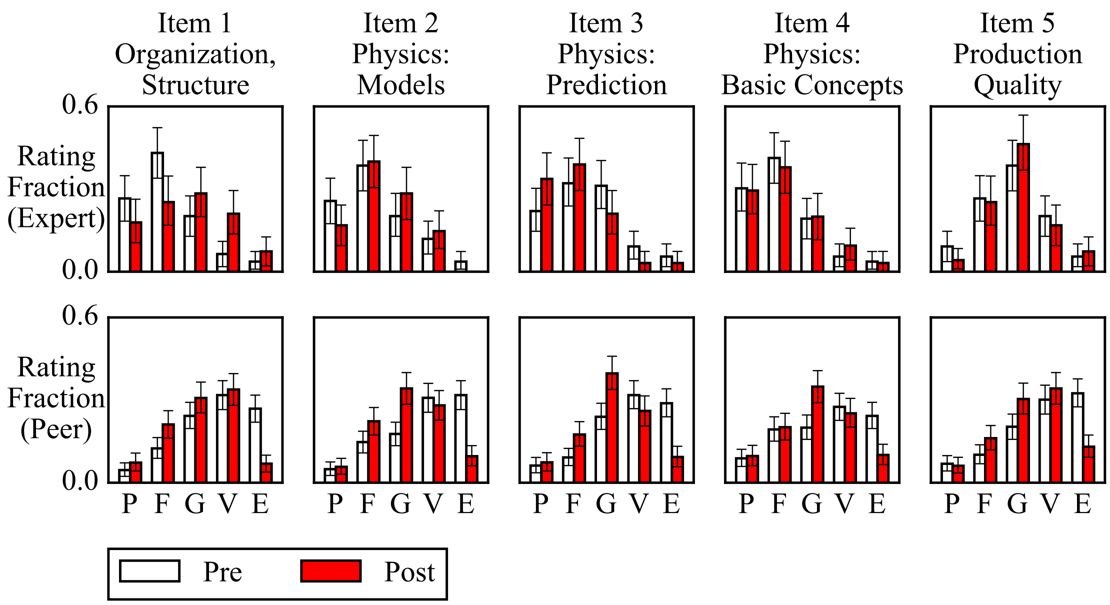


Figure 4.4: Peer ratings for physics-content rubric items (2, 3, and 4) show a greater pre/post decline than peer ratings for non-physics-content rubric items (1 and 5). Consistent among all rubric items was a large decline in peer "Excellent" ratings. Expert ratings were relatively stable across rubric items, with the exception of item 1 (Organization & Structure). Figure shows 2820 peer ratings of a sample of 204 physics lab report videos selected from the first and last labs of the semester, along with expert ratings for each of those videos, sorted by the rubric item to which each rating was assigned. Error bars show 95% confidence intervals. See Table 4.2 for a tabulation of comparisons.

Table 4.2: Results of comparisons between peer and expert ratings of two groups of physics lab report videos (204 videos, 3 peers per video, 5 ratings per peer per video). Items 2, 3, and 4 assess physics content; items 1 and 5 do not. Distributions are compared with a KS 2-sample test, with results reported by the KS test statistic D.[41] D_{peer} compares the "pre" peer ratings with the "post" peer ratings, and likewise D_{expert}. D_{pre} compares the "pre" peer ratings with the "pre" expert ratings, and likewise D_{post}. $D \in \mathbb{R}$, $0 \leq D \leq 1$. D (which does not measure central tendency) is not equivalent to Cohen's d, but D does provide a consistent way to compare whether one pair of distributions is "more different" or "more similar" than another pair: $D = 1$ when two distributions do not overlap at all (maximum difference), $D = 0$ for identical distributions (maximum similarity), D varies continuously in-between.

	Item 1	Item 2	Item 3	Item 4	Item 5
D_{peer}	0.20	0.25	0.25	0.17	0.19
D_{expert}	0.27	0.09	0.19	0.04	0.06
D_{pre}	0.53	0.47	0.46	0.43	0.37
D_{post}	0.15	0.29	0.48	0.38	0.23

As measured by the KS test statistic, the peer and expert distributions became more similar to each other over the course of the semester in items 2 (physics models, $D_{\text{pre}} = 0.47$, $D_{\text{post}} = 0.29$) and 4 (general physics concepts, $D_{\text{pre}} = 0.43$, $D_{\text{post}} = 0.38$), but not item 3 (prediction discussion, , $D_{\text{pre}} = 0.46$, $D_{\text{post}} = 0.48$).

In all cases, the most common expert rating for physics-content items was "Fair", and the proportion of peer "Excellent" ratings fell over the course of the semester. In all "pre" cases, the most common student rating for physics-content items was "Very Good" or "Excellent", while in all "post" cases the most common student rating was "Good".

Ratings for Non-Physics-Content Rubric Items

Items 1 and 5, the non-physics-content items, asked about the organizational/structural quality of the presentation, and the audio/visual quality of the video *per se*, respectively. For the full text of the rubric, see Appendix A.

The expert ratings distributions for both non-physics-content items buck the trends of the physics-content items. Item 1 (organization/structure) shows a relatively large pre/post change in expert ratings distributions compared to the peer ratings ($D_{\text{expert}} = 0.27$, $D_{\text{peer}} =$

0.20), and the most common expert rating for this item changes from "Fair" to "Good", unlike any of the physics-content items.

Item 5 (production quality) shows a stable expert ratings distribution with a substantial peak at "Good", unlike any other rubric item. Among both non-physics-content items, as measured by the KS test statistic, student ratings distributions became closer to expert ratings distributions than for physics-content items.

The difference between trends over time between physics-content and non-physics-content ratings suggests that a separate measure of students' physics conceptual development may be interesting; we report on such a measure in the next section.

4.3 Traditional Measures of Physics Conceptual Development

Fig. 4.4 shows that student/expert agreement for the physics-related rubric items did not increase as much as that for the non-physics-related items. The judgment involved in rating explanations of physics concepts involves both conceptual understanding and skill at assessing scientific communication, which are not separable through an examination of the rating changes alone. To put these rating changes in context, we examined students' pre and post scores on a standard survey instrument used for assessing physics conceptual understanding, the Force and Motion Conceptual Evaluation (FMCE) [69], and also compared the students in our reform course (who did peer assessment) with those in a simultaneous control course using the *Matter & Interactions* curriculum (who did not do peer assessment).

FMCE analysis demands special care because the FMCE was not designed to produce a single-score result, but was rather intended to be interpreted by examining students' responses to specific sets of questions to draw inferences about their understanding of particular concepts [69]. The authors of the FMCE encourage researchers to stick to this form of analysis when discussing student understanding; in cases where a high-level, single-score analysis of FMCE performance is desired, however, the authors provide a rubric for scoring

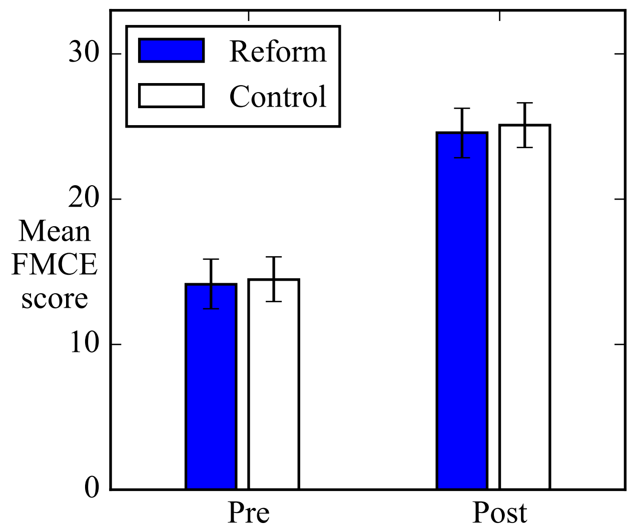


Figure 4.5: Mean matched pre and post FMCE scores for 104 reformed students (the "Your World is Your Lab" curriculum) and 108 control students (the standard Georgia Tech *Matter & Interactions* curriculum) show similar gains in mechanics understanding. These scores indicate that the introduction our laboratory activities and the switch from lecture-based to "blended" pedagogy neither increased nor decreased students' gains in understanding of the mechanics concepts covered by the FMCE. FMCE scores calculated on a 33-point scale. Error bars show bootstrapped 95% confidence intervals.)

the 43-item FMCE on a 33-point scale [69]. We use this rubric here.

Our FMCE results show statistically indistinguishable mean scores for both the control and reform courses. There was no significant difference between the mean pre scores of each group ($\text{Control}_{\text{pre}} = 14.46$, $\text{Reform}_{\text{pre}} = 14.14$), and no significant difference between the mean post scores of each group ($\text{Control}_{\text{post}} = 25.09$, $\text{Reform}_{\text{post}} = 24.57$).

It is customary to report the impact of instructional reforms in terms of the average percentage gain of FMCE scores, $G = (O - I) * 100\%$, where I is the mean fractional incoming ("pre") score and O is the mean fractional outgoing ("post") score for a group of students. It is also customary to report normalized gain $g = (O - I)/(1 - I)$, representing the gain achieved by a group of students as a fraction of *total possible gain* for a given incoming mean I. Normalized gain is sometimes known as "Hake gain" in PER literature, referring to the statistic reported in Hake's classic study of Force Conceptual Inventory scores in interactive-engagement courses [5].

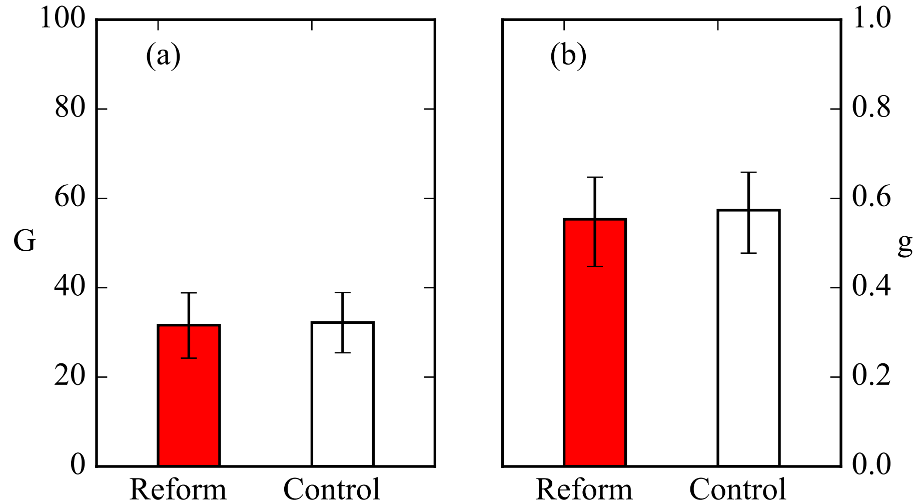


Figure 4.6: Gain (a) and normalized gain (b) FMCE scores are similar for 104 reformed students (the "Your World is Your Lab" curriculum) and 108 control students (the standard Georgia Tech *Matter & Interactions* curriculum. Error bars show bootstrapped 95% confidence intervals.)

As with mean scores, the FMCE gain was statistically indistinguishable between the control and reform courses, which implies that our reform course (with peer assessment) was neither more nor less effective in assisting students' physics conceptual development than the control course. The coincidence of improved student/expert agreement in lab report video assessments and improved FMCE scores admits the possibility (but does not prove) that the improvement in student/expert agreement on physics-related rubric items does indeed reflect an improvement in students' physics conceptual understanding.

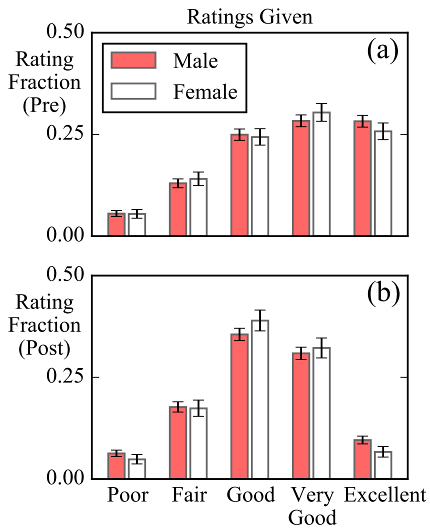
4.4 Ratings by Subpopulation

We used a technique similar to that used in Section 4.2.3 to investigate the possible relationships between rating behavior and potentially relevant subpopulation characteristics such as demographic traits and academic achievement. In this section, we examine the pre and post rating distributions of students divided into subgroups by gender, and by incoming grade point average (GPA). The division of the student population into subgroups introduces the possibility of a difference between the ratings *given by* a subgroup and those *received by* that same subgroup; here, we examine both cases for both subgroups. Since these ratings were not matched to expert ratings, we included the entire set of student ratings in our analysis.

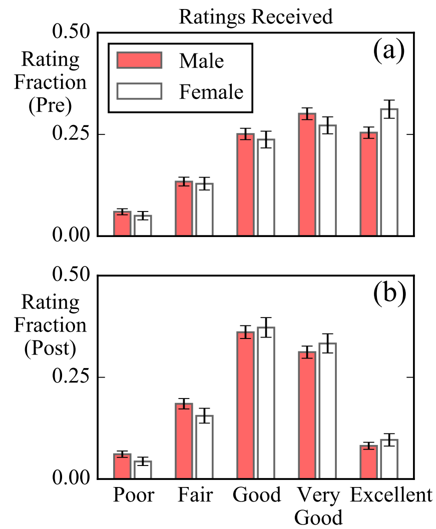
4.4.1 Ratings by Gender

Figure 4.7 compares rating distributions given and received by subgroups of students divided by gender. The change over time in rating distributions, in all cases, reflects that of the changes exhibited in the student population overall (see Figure 4.3); male and female students both receive and give high ratings at the beginning of the semester, with a shift toward the middle of the rating scale at the end of the semester in all cases. However, while a KS test reveals no significant difference between the distributions of ratings *given by* male students and female students ($p_{\text{given, pre}} = 0.466$, $p_{\text{given, post}} = 0.338$ respectively), there *is* a significant difference between the distributions of ratings *received by* male students and female students, both at the beginning and the end of the semester ($p_{\text{received, pre}} = 0.001$, $p_{\text{received, post}} = 0.016$ respectively). Examination of the pre ratings distributions shows that female students began the semester receiving more excellent ratings than any other rating (with "excellent" constituting 31% of ratings received), compared to male students whose received rating distributions peaked at very good (with "excellent" constituting 25% of ratings received).

Post ratings distributions indicate that female students received slightly more excellent (+1.5 percentage points), very good (+2.1pp), and good (+1.1pp) ratings than male students at the end of the semester, and slightly fewer fair (-3.0pp) and poor (-1.8pp) ratings. Altogether, pre and post comparisons of ratings received across genders showed statistically significant differences, but effect sizes were small ($D_{\text{pre}} = 0.06$ and $D_{\text{post}} = 0.05$) relative to the other peer/expert pre/post comparisons reported previously in this chapter (see Table 4.2). Gender therefore appears a statistically significant factor in ratings received, but much



(a) Peer ratings given, by gender.



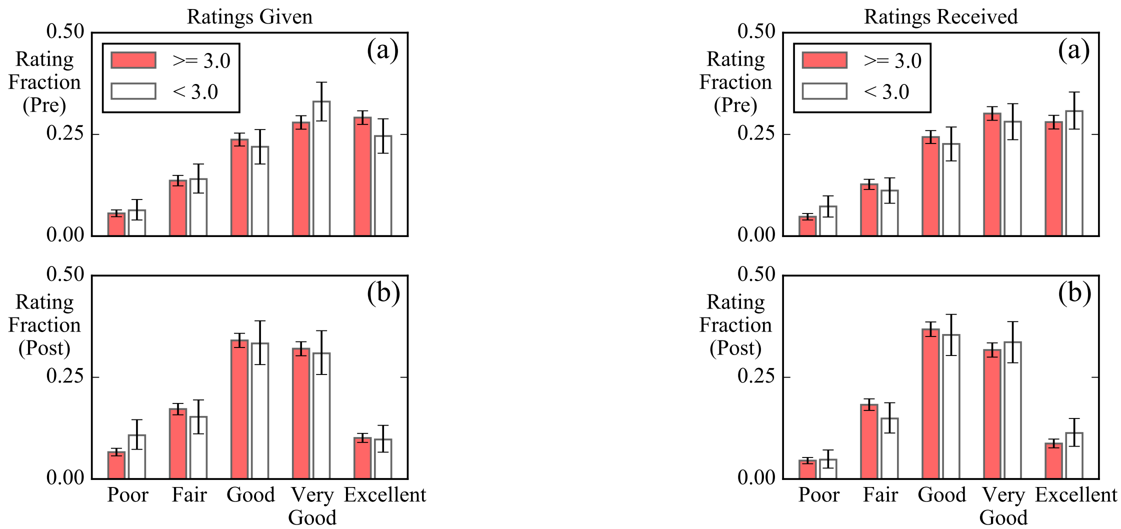
(b) Peer ratings received, by gender.

Figure 4.7: Ratings given by and received by male and female students (N=7445 ratings). Overall, changes over time in these ratings distributions reflect those of the changes over time in the overall student population (see Figure 4.3). A KS test reveals no significant difference between pre and post ratings given by male and female students ($p_{\text{given, pre}} = 0.466$, $p_{\text{given, post}} = 0.338$ respectively), but does reveal that pre and post ratings received by male and female students are significantly different ($p_{\text{received, pre}} = 0.001$, $p_{\text{received, post}} = 0.016$ respectively). Ratings received by female students are slightly but significantly higher than those received by male students, as evidenced by a shift in rating distributions toward the higher end of the rating scale. Pre and post comparisons of ratings-received distributions divided by gender show KS effect sizes of $D_{\text{pre}} = 0.06$ and $D_{\text{post}} = 0.05$, respectively, indicating differences smaller than almost all comparisons between peer and expert ratings distributions in this study (see Table 4.2).

smaller in effect than peer/expert status or the effect of change over time.

4.4.2 Ratings by Incoming GPA

Figure 4.8b compares rating distributions given and received by subgroups of students divided by incoming GPA above and below 3.0 (N=5512 ratings; differences in N values with respect to Figure 4.7 stem from incomplete grade records). In all cases, the change over time in rating distributions reflects that of the changes in the student population as a whole, and a KS test reveals no significant difference between pre and post distributions of ratings given ($p_{\text{given, pre}} = 0.965$, $p_{\text{given, post}} = 0.569$ respectively) or of ratings received



(a) Peer ratings given, by GPA.

(b) Peer ratings received, by GPA.

Figure 4.8: Ratings given by and received by students with an incoming GPA of 3.0 or greater, and students with an incoming GPA of less than 3.0. Overall, changes over time in these ratings distributions reflect those of the changes over time in the overall student population (see Figure 4.3). No significant difference was found between high- and low-GPA students' rating distributions, neither for ratings given ($p_{\text{given, pre}} = 0.965$, $p_{\text{given, post}} = 0.569$) nor ratings received ($p_{\text{received, pre}} = 0.493$, $p_{\text{received, post}} = 0.749$).

($p_{\text{received, pre}} = 0.493$, $p_{\text{received, post}} = 0.749$ respectively). Incoming GPA, therefore, did not appear to exert an influence on ratings distributions given or received, or on changes over time in those ratings distributions.

4.5 Generalizability and Repeatability

In the following sections, we perform analyses on student assessments of training videos and of peer videos. The training videos are the 10 lab report videos selected by instructors from a 2013 offering of the course to serve as calibration and practice videos (see Chapter 2); the peer videos are a sample of 204 lab report videos selected from the beginning and end of the 2014 offering to serve as a research sample. Our analysis of student assessments of the training videos in particular merits special consideration, since this analysis has several significant drawbacks, particularly the relatively small number of videos and the biased process used to select them. The training videos were not drawn randomly from the set of previous semesters' lab report videos, nor was the analysis controlled *ex post facto* to account for the selection process; rather, the videos were selected individually by instructors to provide a reasonably comprehensive set of examples spanning the range of quality and physics content which instructors expected students to encounter in their peers' videos. The resulting set of videos was therefore not likely to be representative of the larger population of peer videos, and the (extremely subjective) decisions involved in the selection process are probably not recoverable.

Nevertheless, we deemed student assessments of these videos worthy of investigation because these videos were consistent across all semesters, and so provided a means by which to compare rating trends from different semesters and different classroom contexts. Even if the the analysis of training videos could not necessarily be said to apply to the peer videos, we reasoned, any evidence of cross-semester consistency of training video rating trends would provide support for an overall cross-semester consistency in students' experience with our peer assessment system.

We present in this section our analysis of students' training video assessments in the Spring 2014 on-campus course, the 2016 online course, and the 2017 online course (see Chapter 2 for a full description of classroom settings). We find a cross-semester consistency in the pre and post distributions of training video ratings, providing evidence that student rating behavior is consistent across different classroom contexts and throughout time.

The students we interviewed in 2016 and 2017 took the course in a different context than did our 2014 students, even though the curriculum and peer assessment system were unchanged. The sole 2016 section comprised 21 students and was conducted entirely online, likewise the 2017 section with 28 students. The 2014 sections, on the other hand, comprised 338 students and had online and on-campus components.

Nevertheless, the peer rating distributions of the 10 pre/post videos rated in common by all online and on-campus peers in all years show a very high degree of similarity (see

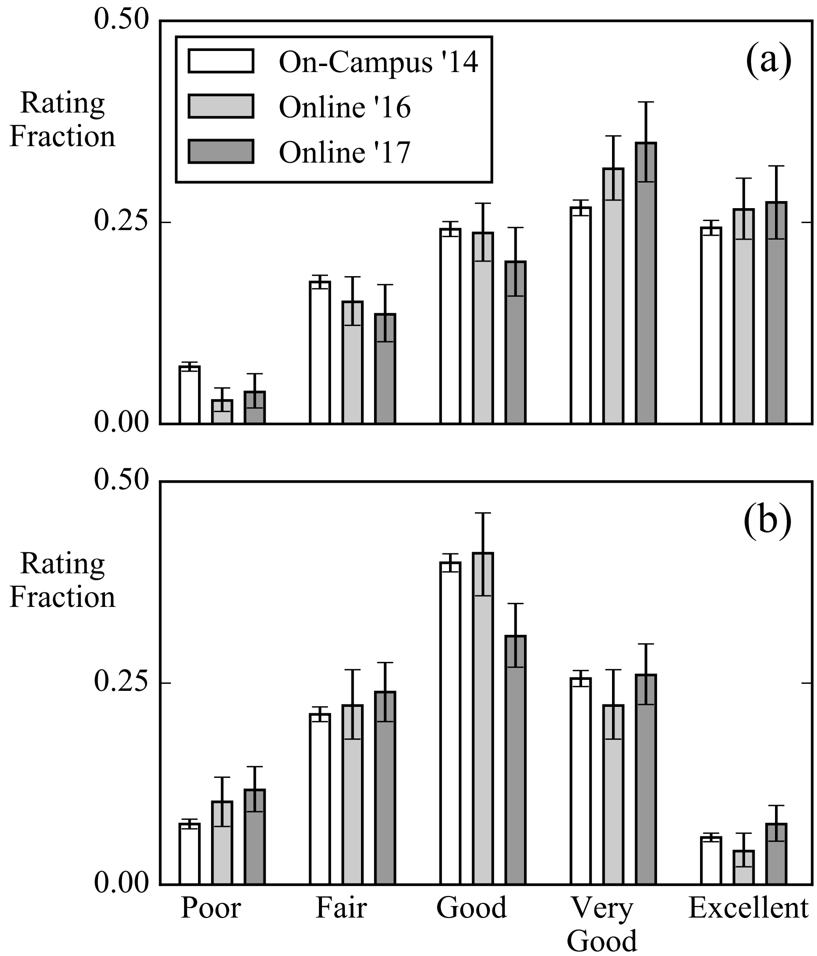


Figure 4.9: Peer rating distributions in two different classroom contexts—on-campus and online—show the same pre (a) and post (b) trends. The on-campus offering was conducted in 2014, and comprised 338 enrolled Georgia Tech students. The 2016 and 2017 offerings were conducted online, and comprised 21 and 28 enrolled Georgia Tech students, respectively. Peer ratings of 10 videos evaluated by students in both course offerings are shown. The similarity of the distributions indicates good overall repeatability, and suggests that the students in both classroom contexts are generally comparable with respect to their peer evaluation behavior. Error bars show 95% confidence intervals.

Fig. 4.9), and resemble the trends we discuss later in the on-campus peer ratings (compare Fig. 4.9 to Fig. 4.3). We therefore believe that the experiences of all cohorts with respect to peer evaluation are comparable, and that the interviews of the online students can yield broadly applicable insights into the peer evaluation process among all cohorts.

4.6 Summary

Quantitative analysis of student ratings of training videos revealed a substantial shift away from the high end of the ratings scale and toward the center of the rating scale, with no substantial change at the low end of the rating scale. This trend was replicated among a random selection of peer videos, and held across multiple administrations of our course. Expert rating distributions, by contrast, remained relatively stable across the semester. We take the change in student ratings as compared with the stability of expert ratings as evidence that student attitudes toward the rating scale, toward their peers' video content, or both, changed over the semester in such a way as to reduce high ratings and promote middling ratings.

Gender appears to exert an influence on ratings; female students *received* slightly higher ratings than male students at the beginning and at the end of the course, though both male and female students *gave* statistically identical distributions of ratings at both times. In all cases, though, the distributions of ratings given and received by male and female students exhibited the same change over time as that of the overall student population; ratings shifted substantially away from the high end of the rating scale and toward the center, and did not change significantly at the low end. Effect sizes of comparisons between male and female ratings-received distributions were also smaller than the effect sizes of peer/expert comparisons or the effect size of changes over time.

We investigate the nature and causes of this broader change in student rating distributions through student interviews in Chapter 5, and correlate this change to certain features of the lab report videos in Chapter 6. In Chapter 7, we summarize our major findings and propose a future research agenda.

CHAPTER 5
STUDENT ATTITUDINAL CHANGES OVER TIME

In this chapter, we report our qualitative investigation of student engagement with peer assessment over time, aiming to expand our understanding of both the pedagogical and logistical implications of peer assessment. As mentioned in Section 2.8, qualitative research aims to answer questions about the subjective experience of the topic at hand; with a qualitative understanding of students' engagement with peer assessment, we can get insight into the attitudes and concepts which students develop (a pedagogical concern) and learn how they might come to adopt accuracy-improving practices (a logistical concern). We conduct our qualitative investigation into student engagement with peer assessment by interviewing students and asking them to narrate their experiences with our course and our peer assessment system.

Specifically, these interviews were intended to explore the trends we discovered in our quantitative work. In Chapter 4, we identified two consistent trends over time among the students who took our course:

- 1.At the end of the semester, students were substantially less likely to assign excellent ratings to each other in their peer assessments, but just as likely to assign poor ratings, resulting in a shift of student ratings from the high end to the central part of the rating scale.

- 2.Student-expert agreement increased overall. This improvement was greater on the rubric items that related to rhetoric and video production than on the items relating to physics content.

Both trends were statistically significant and had an effect size large enough to merit explanation. This change in students' rating behavior was, presumably, the result of some evolution in their views, attitudes, conceptual understanding, and experiences, all of which should be amenable to exploration through qualitative research methods. This chapter focuses on qualitative research into changes in student attitudes and practices over time carried out through think-aloud and semi-structured interviews [4]. This research is intended to supplement our understanding of the trends discovered in Chapter 4, and inform the future development of a model of student engagement with peer assessment.

5.1 Interview Protocol

We conducted student of students who took the 2016 and 2017 SOUP to inform our quantitative analysis with an investigation of student attitudes and reasoning. We interviewed 2016 students after the fourth lab of the course, and interviewed the 2017 students once after the first lab and again after the fourth. In all cases, we administered the interviews online over Google Hangouts, and recorded video of each interview. Each interview was conducted by the same researcher and lasted roughly an hour. The interviews in 2016 began with a think-aloud exercise [70] where the student performed a warm-up problem and then watched and assessed a lab video in the same manner as during the course. The think-aloud portion of the study was immediately followed by a semi-structured interview with followup questions about the assessment exercise, the student's experience in the SOUP, and their attitudes toward assessment, their peers, and video production.

The 2017 interviews were conducted in much the same way as the 2016 interviews, aside from the fact that the 2017 interviews were conducted twice—pre and post—for each interviewee. The main way in which the 2017 interviews differed from the 2016 interviews was by the addition of a second video assessment task for the interviewee; not an assessment of a full lab report video, but of a silent video containing a brief clip of a tennis ball being pushed and rolling across a floor, followed by a Tracker analysis of that same video clip. The video clip was selected from among the 2014 peer videos to be representative of the observational data presented by students in their Lab 1 lab report videos on constant velocity motion (very often balls, jars, or bottles rolling across tables or floors). At the end of the 2017 interviews, students were asked to watch the video, describe the physics of the rolling ball and the shape of the Tracker position-vs.-time plot, and then asked followup questions intending to probe their understanding of friction, rolling, and the summing of forces to produce a net force. The intent of this part of the interview was to explore the way in which students produce explanations of physics concepts at short notice in an unrehearsed manner, how these explanations compare with the formal, rehearsed explanations presented in their lab report videos, and to see how these explanations change over time as students gain experience with peer assessment and develop physics conceptual expertise.

During the lab video assessment, the student's computer screen was recorded, capturing the lab video itself and everything the student typed. At all other times, video of the student's face was recorded. Transcripts were made for each interview (see Appendix F)

5.2 Initial Findings

In these interviews, we explored the attitudes expressed by students about the peer assessment process, the laboratory exercises, and instructor expectations, and encouraged students to reflect on how their attitudes might have changed over time. Our 2017 interviewees, for the most part, expressed the same major themes as the 2016 interviewees. Not every interviewee strictly repeated these themes, of course, and they also expressed different attitudes and observations on a variety of other topics, but the overall consistency of theme between 2017 and 2016 interviews (and the similarity of rating distributions, see Figure 4.3) suggest that students' experience with the peer assessment system is comparable across time and course context.

These interviews were intended to be exploratory—e.g., suggestive of new research questions and lines of inquiry, not intended to confirm or reject specific preconceived hypotheses. We aimed to use this exploration to inform our investigation into the final assessment-related research question identified in Chapter 1: *Why do student assessment practices change in the way they do? What changes in student attitudes and cognition effect these behavioral changes?* This exploration alone yielded no complete answer to this research question for the reason that retrospective accounts of changes in one's own reasoning and attitudes cannot generally be assumed accurate [71]. Instead, these collected interviews served to build a catalog of students' expressed perspectives on and experiences with peer assessment. To the degree that themes expressed in these interviews correspond with or contradict observed changes in behavior, they can potentially be (in)validated with further research and so used to inform the development of a model of student engagement with peer assessment. The following sections of this chapter describe two such themes. A potential research agenda for developing a model of student engagement with peer assessment is described in Chapter 7.

5.3 Qualitative Themes

We found two recurring, related themes among these student interviews which appeared to be relevant to the trends in student rating behavior described in Chapter 4. The first theme was students expressing the belief that experts had critical or harsh attitudes toward grading, while students' own attitudes were "nice", especially toward the beginning of the semester. The second was that low/bad ratings required a higher level of confidence and expertise on the part of the rater than did higher ratings, and that the instructors' greater experience and physics content knowledge endowed them with more of this confidence and expertise than students.

Here we summarize these students' expressions of these two themes and discuss what they might have to say about the observed trends in student rating behavior. All quotes are edited to remove disfluencies. For full transcripts of all interviews and demographics of the interviewees, see Appendix F.

5.3.1 Theme: Being 'Nice' or 'Harsh'

Finally, some interviewed students explained their own "nice" attitudes by describing a feeling of identification with their peers—as one student put it, the author of any given peer video "was in the same position [she] was", and they were "all trying to get a good grade in this class". Other students expressed camaraderie in different ways, such as "we're all at the same level", or by framing the idea in terms of mutual respect for effort; one student said it was important to take grading seriously "because they took time just like I took time to make my video".

Interviewed students who used our peer review system consistently reported a belief that expert/instructor grades of the training videos were more negative or "harsher" than their own, and that this had to do with the relative difference between themselves and the instructors in experience and physics content knowledge. Students said that it was usually easy to tell when a video deserved a poor rating—often because a particular section was completely missing—but not as easy to tell when a video merited a rating somewhere in the middle of the rating scale. Students often reported feeling predisposed to give their peers higher grades, with some students independently describing this predisposition in terms of giving their peers the "benefit of the doubt" or starting from 100 and deducting points when merited. One interviewee during the pre interview touched on all these themes concisely:

> Interviewer: Can you try to characterize the difference... between how the experts assign ratings and how you assign ratings to [the interview video]?
>
> Student A: Like I said, I'm not a hundred percent sure why [the experts] assigned very good [for this item]. I do tend to write as much as I can in terms of like, "oh, she didn't do this", "she did this" but I guess I tend to focus on the positives. Whereas they've they've done it probably hundreds of times... I feel like I just haven't experienced as much, like, content as they have, so they would know what would separate very good from an excellent, and... good

> from very good, and fair... but I guess I do tend to focus more on what they did positively, and then consider what they did negatively after that... I guess, like, what I'm saying is like everyone starts off well for me, but once they do bad—once they do inaccurate things, once they have errors, it goes down from there.

Corroboration with Quantitative Evidence

Of the ten students interviewed, eight reported feeling that their agreement with the expert scores on the training videos improved with time and/or that they were more critical or willing to give lower ratings at the end of the semester than at the beginning. Only one student (Student B), however, actually reported becoming more comfortable giving "poor" ratings *per se*. Retrospective accounts like these alone do not constitute conclusive evidence of actual cognitive or affective changes; however, in this case, the students' accounts so closely replicate the trend toward greater quantitative student/expert agreement that they might very well have explanatory power. As discussed in Chapter 4, student-expert rating agreement improved over the course of the semester, involving a substantial reduction in "excellent" ratings but no significant increase in "poor ratings". This trend was recapitulated in the ratings distributions of three different course offerings (See Fig. 4.9), and, in these interviews, a similar trend was given voice by students describing their own experiences in the course.

As described in the preceding section, students attributed these behavioral and attitudinal changes to a variety of factors, including feelings of camaraderie and reciprocity with peers, confusion with the central ratings on the rating scale, and a rating strategy which affords peers the benefit of the doubt. Further research may more rigorously identify and confirm the specific cognitive or social developments underlying these changes; the close correspondence between students' accounts and observed trends suggest that these developments operate at least partially on the level of the students' conscious awareness.

5.3.2 Theme: Expertise and Criticality

Interviewed students often attributed the differences between peer and expert attitudes to the greater physics knowledge possessed by the experts. With respect to the relative harshness of expert ratings, interviewed students described the experts as being able to attend to and identify smaller and subtler "physics errors" than students, since the experts knew more about physics. Students described themselves as not possessing the knowledge necessary to notice these errors, and so they gave higher ratings to their peers' videos than they would have if they had been able to notice the errors. Follow-up questions revealed the students' usage of "physics error" to mean something like a misconception, *i.e.* a flawed or undeveloped expression of a physics concept, rather than a mathematical or procedural error *per se*.

One student made a particularly clear expression of her own inexperience relative to the instructors, in the context of describing the difference between her ratings and the instructors' ratings:

> Student 2: I guess... the difference between [our ways of] grading comes down to like, I'm a student, I don't know as much as a professor does. And a professor has encountered a lot more, like, training, and a lot more experience to actually grade someone else's work. So, they're a bit more meticulous. And I mean, a professor is teaching a subject so that another person learns the subject. I—I'm learning it, so I'm not as meticulous as someone else, because I'm also trying to grasp the concepts.

One student described her change in rating behavior over the semester in terms of a re-evaluation of her own inclination toward giving higher ratings, while mentioning an evolution in her willingness to be critical:

> Interviewer: You're describing the experts having a different set of knowledge about physics: more knowledge about physics, a better ability to attend to

physics details. But overall, if you had to describe it generally, what do you think is the difference between how the experts and how you evaluate these videos?

Student 1: I think now, I would say I'm a lot more accurate, I guess, in terms of like comparing [my ratings with] the experts. I think the first time I did [an assessment], I think I was like being very nice, and I gave a lot of excellents. And then I realized, like, "Oh, that was like not what it was at all!" So I think over time I've realized to be more critical toward the video, I guess. Cuz at first I was like "Oh, yeah, that was good, that was good!" And then like... as I've gone on, I've like gotten more used to being very critical towards like certain things that people say... I guess, like, if people say something wrong, it's not—I think in the beginning, I was like, "Oh, well, maybe they just like said the wrong thing." But like... they said the wrong thing, so I should mark them down.

Interviewer: So at the beginning you said you thought—at the beginning of the semester you, uh, felt like you were being nice to your fellow students. Why did you feel like being nice to your fellow students?

Student 1: Well, I also think I was like... not sure how harsh to be. [*laughs*] I guess!

Often, students would attribute this inclination toward giving higher grades as stemming from a feeling of reciprocity—a recognition that the other students seemed to have put in just as much effort to making their videos as the interviewee did, and therefore should be rewarded for their effort in the same way the interviewee felt they themselves should be. Another interviewee, during her post interview, linked this reciprocal feeling to her disinclination to give "poor" ratings and the "qualifications" required to do so:

Interviewer: How did [rating the interview video] compare to the rating experience you had at the beginning of the semester?

Student B: I think at the beginning of the semester... I dunno, it's awkward to rate people's video when you just did the same thing. Like, I just felt bad. But through the semester I got more comfortable with it, and it didn't really bother me as much, like, if I knew someone definitely didn't have an intro I would give it a poor or whatever. Or like if they definitely had a big physics error, I wouldn't be as uncomfortable giving a poor. The first time I did it, I don't think I gave anybody a poor. Interviewer: Mm-hmmm.

Student B: But, as it went on, I got more comfortable doing it.

Interviewer: What is it about needing to be comfortable that's so important for giving poor ratings?

Student B: It just makes me feel bad, because I'm sure they put work into it too, you know. Like, I'm not a physics professor, so I don't necessarily feel like I have a hundred percent qualifications, like if I give someone who worked on something a poor. But then... I just tried to stay truer to the rubric, and [base my judgments] on that... I don't know. Because I didn't want anyone to give me a poor, so I didn't want to give anybody a poor.

Interviewer: So you said you're not a physics professor, you didn't have the qualifications to give a poor. Could you tell me a little bit about what you think those qualifications are?

Student B: Well if you're a teacher, then your job is to grade. And I'm just a peer, so like... I wouldn't have any trouble if it was just peer feedback, but knowing what I say determines their grade, it's just like, it put a lot of pressure on me, you know? ...it's not like [the video author] would know [it was me giving the ratings], so it's not necessarily I didn't want to hurt their feelings, but I don't know... I don't know exactly what their thought process were for something they said or didn't say. Like, maybe they KNEW that but they just didn't say it so... I don't know. I just give too much benefit of the doubt, I guess.

Interviewer: And your attitude toward that changed a bit throughout the semester? Would you say that's fair to say?

Student B: I mean, it didn't necessarily change, but like... I understood the rubric more.

Corroboration with Quantitative Evidence

Interviewees often expressed a relationship between experts' relatively lower ratings and their expertise: born of their years-long study of physics itself, their extensive experience in assessing lab report videos, or both. Some interviewees reported that they were able to develop what they considered to be a more expert-like attention to detail, conceptual understanding, and/or awareness of physics errors during the course, and that this was related to their own improved accuracy in rating. Here, the relationship between this theme of students' retrospective accounts and other quantitative data is more elusive than for the first theme, described above. Likely the strongest conclusion we can draw from the work covered in this section is that students' claims of a relationship between their own conceptual development and their improvement in rating accuracy is *not contradicted by* our quantitative research, given that an improvement in rating accuracy (See Fig. 4.4) and a gain in measured conceptual competency (See Fig. 4.6) were coincident.

5.3.3 Other Themes and Behaviors

The exceptions to the above themes were also instructive. One student remarked that TAs—not the course instructors—tended to give *higher* ratings than peers, which suggested to her that students tended to be "nitpickier" than TAs. More than one student expressed frustration at the apparent arbitrariness of instructor ratings, the apparent mismatch between instructor ratings and the valence of their comments, and the lack of specific guidance for intermediate ratings (fair, good, very good). Yet another student was somehow able to acquire an earlier, undistributed draft rubric which the instructors used internally during the rubric development process—this draft rubric included specific criteria for each rubric item and rating on the rating scale. While we ultimately discarded this draft after preliminary inquiry suggested that its exhaustive content might be overwhelming for students, nevertheless this one student claimed to find it helpful during the peer assessment process, and credited the draft rubric with her improvement in expert agreement.

Also notable is a particular behavior shown by some students; during the think-aloud video assessment, some students spent time revising their ratings before declaring they were done with their assessment. During this time, these students either kept their ratings the same or revised them downward—not once was an interviewee observed to revise their ratings upward.

5.4 Summary

Students interviewed about peer assessment expressed a wide variety of retrospective accounts of their behavioral, conceptual, and attitudinal changes. While retrospective accounts alone do not constitute evidence for the actual influence of the themes expressed in the accounts, two such themes appeared to offer a plausible relationship with the rating behavior changes discussed in Chapter 4 and were either corroborated or at least not contradicted by our existing quantitative findings. The first theme related to students' characterization of their own rating behavior as "nice" and the instructors' as "harsh" or more "critical", and the second theme was the expression of students' belief that experts' criticality was born of physics expertise and long-term experience with rating lab report videos.

In the following chapters, we explore students' explanations of physics concepts in their lab report videos and how they change over time (Chapter 6) and, finally, we produce a research agenda for future efforts toward a model of student engagement with peer assessment (Chapter 7).

CHAPTER 6

STUDENT COMMUNICATION CHANGES OVER TIME

Having identified a quantitative trend toward greater student-expert agreement in Chapter 4, and having surveyed students' retrospective accounts of their development as peer raters in Chapter 5, we describe here our analysis of the changes over time associated with content of students' lab report videos. Our research goal in performing this analysis was, broadly, to explore how changes over time in the content of the videos which students produced influenced changes in the ratings which other students assigned.

The work in this chapter covers our efforts to answer the specific pedagogy-related research question, *Which features of students' explanations of the laboratory process do students appear to attend to when assessing videos, and how does this change over time?* As with the work covered in Chapter 5, this work was largely exploratory in nature; as with the work covered in Chapters 3 and 4, our analysis was highly statistical. This work therefore combines elements of quantitative and qualitative research.

We categorized students' lab report videos based on the type and quality of the physics explanations they included in them, and sought to find a relationship between these categories and the trends in student rating behavior we identified earlier. We discovered evidence that, as time went on, the length and sophistication of physics explanations in lab report videos (though not necessarily their correctness) became significantly correlated with differences in students' assessments.

6.1 Classifying Communicative Content

To establish a relationship between student ratings of videos and the communicative practices their peers exhibited in those videos ("communicative content"), we had to develop a scheme for classifying those practices. Our first task in developing this scheme was to identify portions of the first and last lab report videos which were commensurable across the pre/post division. This was not a trivial task; as described in Chapter 2, the first and last labs involved different physical systems and required students to use different physical concepts and practices. In particular:

- Lab 1 required each student to observe an object moving at a constant velocity, either by finding such an object in their environment (*e.g.*, slow-moving traffic) or by contriving a constant-velocity system through their own efforts (*e.g.*, by rolling a ball or bottle across a table). Students were required to record and analyze a video of this constant-velocity motion in Tracker, [27] produce a simulation of this motion in VPython, [29] compare their observed and simulated results, and explain the connection between these results, Newton's Second Law, and the collection of environmental forces acting on the moving object.

- Lab 4 required each student to analyze a single instructor-provided video in Tracker. This video depicted a mass swinging in a vertical plane, attached to a spring that was affixed at the other end to a laboratory workbench Students were required to produce a simulation of this motion *and* of the changes in kinetic and potential energy of the spring-mass-Earth system in VPython, compare their observed and simulated results, and explain the connection between these results, Newton's Second Law, the Energy Principle, and the various interactions between the parts of the system.

6.1.1 The Prediction vs. Observation Discussion

For our analysis of these lab videos, we elected to examine one portion in particular that we deemed relatively consistent between the first and last labs; despite the differences between these labs, students made a *comparison between observed and predicted results, and offered an explanation of that comparison* in each. Making this comparison was a learning goal for both labs, identified explicitly in the rubric. The prediction/observation discussion consistently involved a description of a position-vs.-time plot in both labs, and the explanation of that plot involved factors that could affect motion or the appearance of motion; students were expected to be familiar with these concepts by the time the first lab was assigned. Significantly, the prediction vs. observation in each video also tended to be structurally similar between videos, with students usually starting the discussion by showing the computer simulation on-screen, then showing the superimposed prediction and observation plots, then explaining the comparison, and finishing by announcing the start of the subsequent "What if?..." and "What does it mean?" section.

The prediction vs. observation discussion in each video therefore represented a communicative and conceptual exercise comparable across the pre/post divide, was easily identified when searching through student videos, had a consistent structure among different students, and had a beginning and end set by reasonably objective criteria. For these reasons, we selected it for our first in-depth examination of the relationship between student communicative content and student rating behavior. Subsequently in this chapter, all discussion and analysis pertaining to "video content" or "videos" refers to the portion of each video containing the prediction vs. observation discussion in particular.

A prediction vs. observation passage from a selected lab report video is diagrammed in

6.1.2 Prediction vs. Observation Coding Scheme

We created codes for classifying the prediction vs. observation discussion through the method of exhaustion. We wrote a brief summary of each discussion in the set of lab report videos used previously in our research (see Chapter 4), and continued to generate summaries until we found that each new summary was a repetition of an existing one. Having reached this state of "exhaustion", we used our summaries to generate codes, evaluated the codes for face validity, and iterated the process until we were confident we had a set of codes that was applicable to the entire research set and more or less comprehensive.

During our code development process, we found that students tended to first compare

"The result of this code can be seen in the following video, which produces a model similar to the observed model…

…We can see that when you look at the graph comparing the two lines of the observed and computational models, the computed model overestimates the distance travelled by the ball…

…This is because we assume that there are only two forces acting on the observed model, that cancel each other out. Gravitational force, and the normal force, the force applied from the surface of the ground."

Figure 6.1: Transcription of a representative prediction vs. observation discussion from the first lab of the semester, with corresponding freeze-frames. The visual component of the discussion began with the on-screen presentation of that lab's computational simulation, proceeded to a plot comparing the predicted trajectory to the observed trajectory, and ended with a slide attributing the results of the comparison to various factors. This sequence was typical of most students' prediction vs. observation discussions.

the predicted and observed trajectories by describing the degree to which the trajectories were similar or dissimilar, and then to name and explain the factors to which they attributed the (dis)similarity. It quickly became apparent we would require two separate sets of codes to classify these *comparison* and *attribution* phases. A set of "comparison" codes was constructed to capture the sorts of comparisons that students made between the predicted and observed trajectories, and is enumerated in Appendix C. We mention the comparison codes here for the sake of completeness, but no significant trends involving these codes were found and we will not discuss them further here.

The second set of codes—the "attribution" codes—was constructed to capture the factors to which students attributed the discrepancies (or similarities) they described in their comparisons, is fully enumerated in Appendix D, and yielded our first quantitative connection between video content and student rating behavior.

Attribution Codes

After students presented their prediction vs. observation comparisons in their videos, they then tended to attribute their results (usually, the difference between the predicted and observed trajectories) to a number of specific *factors*. These factors included forces such as friction and drag, environmental factors like the shape or texture of the floor, "human error" of various sorts, and a persistent "force of the push" that set the object in motion, a common introductory mechanics preconception [72]. The full codes for each factor are listed in the first column of Table D.1 and reproduced here:

- a.1: Friction (from surface)
- a.2: Drag (from air)
- a.3: Human Error (in Tracker or experiment)
- a.4: Initial Force

- a.5: Gravity
- a.6: Normal Force
- a.7: Shape/Texture of Object/Surface
- a.8: Path of Object is Unusual/Imperfect/Multidimensional
- a.9: Framerate/Recording Error
- a.10: Other/Unknown Forces or Factors (not specified)
- a.11: Energy Transfer
- a.12: Rounding/Numerical Error
- a.00: Other

Students in the videos we investigated attributed their results to up to 6 factors, but typically just to one or two (the 25% and 75% quantiles, respectively). Students then tended to offer explanations of the role of those factors in producing the results they found. These explanations were of varying completeness and correctness, for which we produced the following four subcodes; some students...

- *.1 simply mentioned the name of that factor in passing.
- *.2 offered a very terse or confusing summary of that factor's effects.
- *.3 offered an explanation with a distinct cause and effect but flawed physics.
- *.4 offered an explanation with a distinct cause and effect and correct physics.

These four codes are listed in full in the second column of Appendix D. For clarity, we include two examples of students' prediction vs. observation comparisons and the attribution codes we assigned to them.

Example A

In the final lab, one student compared their predicted and observed trajectories thusly:

> The two start out the same, which makes sense *because the observed data was used to initialize the model*. But they don't match well beyond that point. *There are many reasons to explain this. First and foremost is that it's difficult to mark the position of the mass accurately and precisely in Tracker. There were also rounding issues that may have contributed, along with the fact that the calculation of T, k, and* L_0[1] *only took into account the oscillation in one direction.*

Above, we indicate with italics the language this student used to *attribute and explain*, rather than to *describe*, the results of their comparison. We assigned attribution codes to this language according to the factors the student mentioned: "many reasons", not all of which were specified (a.10), difficulty using Tracker (a.3), and "rounding issues" (a.12).

The student's final attribution—the fact that certain parameters were calculated with data from an oscillation in one direction only—was unique to this student, and so was assigned a.00 and ignored. Only two instances of the a.00 code were assigned to the $N = 204$ research videos; our set of attribution codes was almost but not quite comprehensive.

Having assigned these three attribution codes, we then evaluated the completeness and soundness of the explanation which the student gave for each attributed factor.

- a.10 Other/Unknown Forces or Factors (unspecified): The student merely mentioned the existence of other factors beyond the two they specified, and so we appended this code with a .1
- a.3 Human Error (in Tracker or experiment): The student went beyond simply naming Tracker error as a relevant factor. They described a specific cause (difficulty

[1]These are parameters calculated by students from observational data for use in their simulations (y-oscillation period, spring constant, and spring rest length, respectively.

marking the mass in Tracker) and associated it with a specific effect (the prediction and observation don't match well after a certain point in time), but the physics is suspect; difficulty marking positions in Tracker would tend to produce random errors or a fixed offset in the observed position, not an error that increases systemically over time. For this reason, we appended this code with a .3

- a.12 Rounding/Numerical Error: The student mentions rounding error only in passing and so we appended this code with a .1

The attribution codes we assigned to this video were therefore a.10.1, a.3.3, and a.12.1. No other attribution codes were assigned to this video.

Example B

Another student, in the first lab, discussed their predicted and observed trajectories thusly:

> We see that the experiment's data line slightly curves down compared to the model's data line. *From this, I can conclude that there is a slight net force from the surroundings to the system... I can assume there is a net force pushing against the object because its velocity is slowing down. These forces include air resistance and friction, possibly other unknown forces.*

Again we indicate the attribution language with italics. This student attributed their findings to forces including air resistance (a.2), friction (a.1), and other unspecified forces (a.10). We evaluated the completeness and soundness of the student's explanation for each of these attributions:

- a.2 Drag (from air): The student identified air resistance (which we deemed synonymous with drag) as contributing to the net force and effecting a downward (*i.e.*, toward $-x$) curve in the observed trajectory's position vs. time plot. A drag force would indeed cause this effect in the system this student constructed. This explanation named a specific cause, tied it to a specific effect, and contained sound physics, so we appended this code with a .4

- a.1 Friction (from surface): The student identified friction as contributing to the net force and effecting a downward curve in the observed trajectory's position vs. time plot. This physics is suspect. Friction plays a somewhat complicated role in rolling without slipping; if the object is rolling on a level surface without slipping while also experiencing a drag force, for example, the surface contact force would need a nonzero component *in the direction of motion* to constrain the object not to slip. Depending on the slope of the surface, object/surface adhesive forces, and the magnitude of the drag force, the "friction" force may point either toward or away from the direction of motion, or may be zero. In this system as specified in the video, friction therefore cannot be said to simply slow the object down, as the student does. For this reason, we appended a .3 to this code.

- a.10 Other/Unknown Forces or Factors (unspecified): The student alluded to other unknown forces without offering any details, so we appended a .1

The attribution codes we assigned to this video were therefore a.2.4, a.3.3, and a.10.1.[2]

6.1.3 Explanation Completeness

An important subgoal of our investigation into the relationship between video content and student ratings was to determine whether students' communicative practices changed over the course of the semester *distinct from* their change in conceptual understanding, and whether these communicative practices had an effect on student rating behavior; to this end, we intuited it was important to disentangle the *completeness* of the students' explanations

[2]We also assigned the code a.E—the "hedged language" code, defined in Appendix D—because the student said the system "possibly" exhibited other forces. This code is irrelevant to the discussion in this chapter, and so is excluded from the body of the text.

(related directly to communicative practice) from the *correctness* of those explanations (related directly to conceptual understanding).

Our "explanation" codes (the appended *.1-*.4 codes, described above and in Appendix D) seemed to satisfy our requirement for encoding completeness (though not necessarily correctness). Codes *.3 and *.4 shared the property of encoding explanations with distinct cause and effect, the conceptual correctness of which could be judged at face value—*i.e.* "complete" explanations. Codes *.1 and *.2 both encoded explanations without cause and effect, but which were too terse or superficial to admit a face-value assessment of the conceptual understanding of the author—"incomplete" explanations. We therefore found it useful to collapse the explanation codes into a single dichotomous "completeness" variable, with a.*.3 & a.*.4 constituting "complete" explanations and a.*.1 & a.*.2 constituting "incomplete" explanations.

To further simplify our analysis of explanation completeness in relation to student ratings, we assigned an overall "complete" or "incomplete" dichotomous code to each video depending on whether the video contained a majority of complete explanations as defined above. Since some video contained a mix of complete and incomplete explanations of different factors as demonstrated in the above examples, we needed to select a criterion for assigning an overall completeness to the whole video. To select this criterion in an informed rather than an arbitrary way, we first examined the ratio of complete to incomplete explanations present within each video (the "completeness ratio", shown in Figure 6.2). We defined the completeness ratio of a video to be the ratio between the number of complete physics explanations in the prediction vs. comparison discussion; *i.e.*, $R = \frac{\mathrm{N_{com}}}{\mathrm{N_{com}}+\mathrm{N_{incom}}}$, $0 \leq R \leq 1$, where R is the completeness ratio of a video, $\mathrm{N_{com}}$ is the number of complete physics explanations in that video's prediction vs. observation discussion, and $\mathrm{N_{incom}}$ is the number of incomplete physics explanations in that prediction vs. observation discussion. Returning to our examples above, the video in Example A had three explanations, two of which were "incomplete" (a.10.1 & a.12.1) and one of which was "complete" (a.3.3), so the

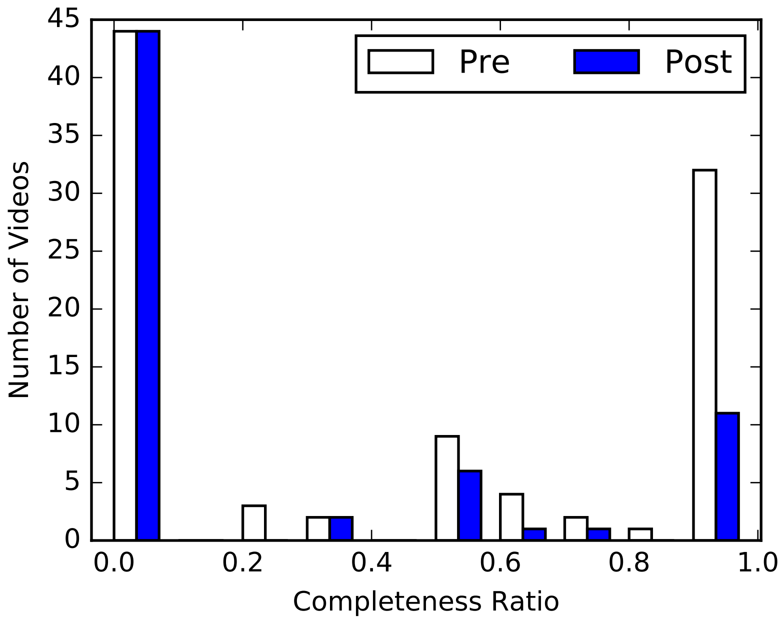


Figure 6.2: Distributions of completeness ratios for 204 randomly-selected student lab report videos from Spring 2014 show that most videos contain either entirely incomplete or entirely complete physics explanation, and that fewer videos with entirely complete explanations show up in the final lab than in the first one. Discrepancies between the number of videos represented in the plot (162) and the total number of selected videos (204) arise from videos with no physics explanations at all being excluded from analysis.

video had a completeness ratio of $R_A = \frac{1}{3} < 0.5$ and was therefore marked "incomplete" overall. Example B also had three examples, but only one was incomplete (a.10.1) and the other two were complete (a.2.4 & a.3.3), so its completeness ratio was $R_B = \frac{2}{3} > 0.5$ and it qualified as "complete" overall.

The bulk of videos both pre and post contained only incomplete explanations, and so had a completeness ratio of 0; many other videos had a completeness ratio of 1 (complete explanations only), though fewer post than pre. A substantial portion of videos had an intermediate completeness ratio between 0 and 1; see Figure 6.2.

We ultimately decided to call videos with a completeness ratio $R > 0.5$ "complete" overall, and ones with a completeness ratio $R \leq 0.5$ "incomplete" overall. When we selected the strictest possible completeness criterion—assigning a whole video "complete" only if it had a completeness ratio of 1—we found no significant difference in the effects described later in this chapter than when we selected a completeness ratio > 0.5. The effects described below only disappeared when we selected the laxest possible completeness criterion (completeness ratio > 0), which we considered to be unreasonably lax, since we did not expect that a typical rater would consider an entire video to contain generally complete explanations if only a small minority of them were actually complete. The effects described below, then, appear to be robust to changes in the overall completeness criterion (above a reasonable minimum), and therefore not merely artifacts of an arbitrary choice of criterion.

6.2 Completeness vs. Student Ratings of Physics Content

Our overall goal in this analysis was to determine whether and what sort of quantitative relationship between existed between student-produced video content and student rating behavior. To narrow down our research focus, we devoted our efforts to answering the specific research question, *Which features of students' explanations of the laboratory process do students appear to attend to when assessing videos, and how does this change over time?* We therefore separated videos into "complete" and "incomplete" categories depending on the overall completeness of the explanations within them, as described above, to see if this feature students' videos had an influence on peers' ratings of physics content during peer assessment, and whether that influence changed over time.

We examined peer ratings of physics-content-related rubric items (Models, Prediction Discussion, and Physics Overall: see Appendix A) of the same set of student videos in Chapter 4, this time sorted by completeness. Our results are shown in Figure 6.3. We found that the pre rating distributions were statistically similar across completeness (Mann-Whitney U: $p = 0.051$), the *incomplete* videos were the ones which exhibited a pre/post change in student rating distributions (Mann-Whitney U: $p << 0.05$, effect size 0.642—*i.e.*, incomplete post ratings are 62.2% likely to be less than incomplete pre ratings) and the complete videos saw no significant pre/post change (Mann-Whitney U: $p = 0.164$). In particular, the post distributions show a stark difference in the percentage of "excellent" ratings, 28% among the complete videos and 8% among the incomplete videos (dichotomous Chi-Squared: $p << 0.05$), while the pre videos saw no significant difference in "excellent" ratings across completeness (dichotomous Chi-Squared: $p = 0.442$). The trend in student ratings we found in Chapter 4 (toward greater peer/expert agreement and away from "excellent" ratings) therefore appears to be stronger among the incomplete videos than among the complete videos.

Performing the same analysis on the non-physics rubric items for the sake of comparison, we found no significant across-completeness difference between the percentage of "excellent" ratings in either the pre (dichotomous Chi-Squared: $p = 0.219$) or post (dichotomous Chi-Squared: $p = 0.143$) distributions. The students therefore appear to distinguish between videos containing complete vs. incomplete physics explanations only when assessing physics-related rubric items, which is what we would expect if the physics-related rubric items and our complete/incomplete coding scheme were both at least minimally valid measures of physics content.

6.2.1 Completeness vs. Expert Ratings of Physics Content

To provide a baseline comparison to the student results above, we examined the expert ratings for the physics-content rubric items for the same videos (again drawing on the expert ratings for the lab report videos from Chapter 4). We do not assert that these physics-content ratings constitute a measure of physics conceptual understanding completely independent of student communicative practice; nor do we expect that the physics content items on the rubric have any particular correspondence with *e.g.* items on a validated conceptual survey like the FMCE [1]. We did, however, wish to provide a baseline for assessing student physics understanding *in the context of producing a video lab report*, which—appealing to the resources framework [73]—could take a different expressed form than understanding in the context of a multiple-choice conceptual survey or a timed exam, and these expert rat-

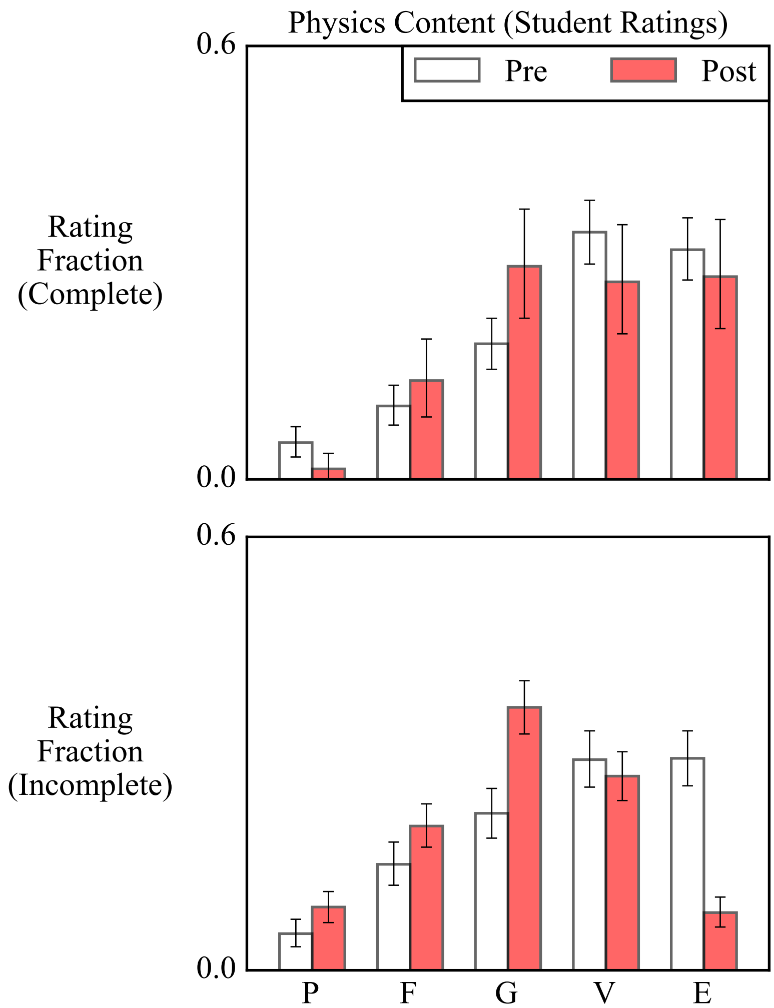


Figure 6.3: Distributions of student physics-item ratings (N $= 1795$) for 204 randomly-selected student lab report videos from Spring 2014. Ratings are separated by pre and post (white and red, respectively) and by whether the video contained complete physics explanations (top) or incomplete physics explanations (bottom). A Mann-Whitney U-test reveals that pre distributions are statistically indistinguishable across completeness ($p = 0.051$), no significant pre/post change in the typical student rating for complete explanations ($p = 0.164$), and a significant decline in typical student ratings for incomplete explanations ($p << 0.05$, effect size 0.642). Discrepancies between the ideal number of physics-item ratings ($N_{ideal} = 3$ physics items per video $\times$ 204 videos $\times$ 3 raters per videos $= 1836$) and the actual number (1795) arise from some videos not containing a prediction vs. observation discussion at all, and being excluded from the analysis. Error bars show bootstrapped 95% confidence intervals.

ings of physics content seemed to be the most direct measure of conceptual understanding in that context available to us.

Using expert ratings as the criteria for assessing expressed physics conceptual understanding also allows us to compare our analysis of communicative content with the rating-behavior analysis of Chapter 4.

Sorting pre and post videos by completeness as before and examining the distribution of expert ratings, we find a different trend than that which we found among the student ratings. Figure 6.4 shows that the expert rating distributions for both “complete” and “incomplete” videos are statistically similar at the beginning of the semester (Mann-Whitney U: $p = 0.911$), but are different at the end of the semester (Mann-Whitney U: $p = 0.002$, effect size 0.647) with a marked decline in the percentage of “poor” ratings among the complete videos (27% to 10%, dichotomous Chi-Squared: $p = 0.048$) but not the incomplete videos (dichotomous Chi-Squared: $p = 0.196$).

6.3 Discussion

Here we reiterate the trends we found in student and expert ratings of complete vs. incomplete videos, and incorporate those trends with our findings from previous chapters. We propose a potential mechanism underlying our observed trends and suggest an avenue for further research.

6.3.1 Student & Expert Rating Trends

We found that students did not appear to distinguish between complete and incomplete explanations when assigning ratings to physics content at the beginning of the semester, but did give higher physics-content ratings to videos with complete explanations at the end of the semester. We observed a different trend among expert ratings of those same videos; while expert ratings were also similar across completeness in the beginning of the semester, their ratings of incomplete videos held steady while their ratings of complete

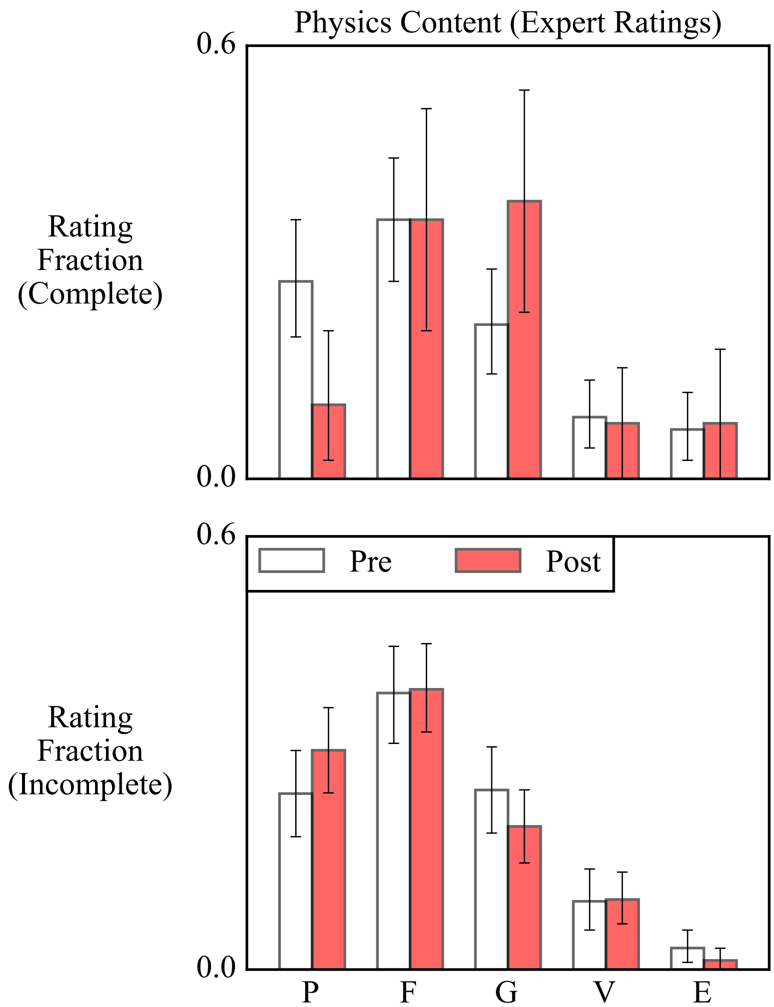


Figure 6.4: Distributions of expert physics-item ratings ($N = 594$) for 204 randomly-selected student lab report videos from Spring 2014. Ratings are separated by pre and post (white and red, respectively) and by whether the video contained complete physics explanations (top) or incomplete physics explanations (bottom). A Mann-Whitney U-test reveals that pre distributions are statistically indistinguishable across completeness ($p = 0.455$), no significant pre/post change in the typical expert rating for incomplete explanations ($p = 0.104$), and a significant rise in typical expert ratings for complete explanations ($p = 0.044$, effect size 0.604). Discrepancies between the ideal number of physics-item ratings ($N_{ideal} = 3$ physics items per video $\times$ 204 videos $= 612$) and the actual number (594) arise from some videos containing no prediction vs. observation discussion and being excluded from the analysis. Error bars show bootstrapped 95% confidence intervals.

videos improved.

The pre/post decline in student ratings among the incomplete videos—particularly the dramatic decline in "excellent" ratings—seems to recapitulate the overall trend in student rating behavior we observed in Chapter 3, and the comparatively stable student ratings of complete videos suggests that completeness (as we defined it) is indeed a relevant factor in the evolution of student rating behavior over time.

What's more, the similarity of students' pre ratings across completeness, the stability of their complete ratings over time, and the decline of their incomplete ratings over time all together corroborate some themes of students' retrospective accounts that we uncovered during our interviews in Chapter 4. Interviewed students commonly reported beginning the semester with an inclination to give their peers high ratings during assessments, and some also reported a persistent tendency to begin assessments with a high grade in mind and to dock points as they encountered errors. Notably, many interviewed students also reported a greater ability to discern "physics errors" in the lab report videos as the semester progressed. If the incompleteness of physics explanations had indeed become more salient as a physics error to student raters as the semester progressed, we would expect the distributions of student ratings of the "incomplete" videos to undergo the changes we observed (*i.e.*, a decline in ratings).

We would also expect student ratings of the "complete" videos to have remained relatively stable as incompleteness became more salient to student raters, as we observed. If the decline in student ratings of incomplete videos were caused by students becoming more likely to dock points for incompleteness, then videos which were not incomplete would be exempt from the decline.

It is unlikely that a single, simple underlying mechanism is responsible for the trends we observe in student ratings among the complete and incomplete videos; nevertheless, the combination of the observed decline in student ratings of incomplete videos, the stability of student ratings of complete videos, and with plausible and apparently corroboratory trends uncovered in student interviews does suggest that the completeness of physics explanations is a relevant factor in the time evolution of student engagement with our peer assessment system.

6.3.2 Student Communicative Trends

The analysis covered in Chapter 5 also allowed us to explore the changes in the student *production* of scientific communication, not just peer *rating* of it. Two quantitative trends combine to suggest that students' practices in constructing formal physics explanations was itself subject to change over the course of the semester. The first such trend is the decline in the number of sampled videos that we classified as "complete", along the stability of the number of "incomplete" videos (See Figure 6.2). The second such trend is the pre/post improvement in expert ratings of videos with complete explanations; taking expert ratings of physics content to be a ground-truth assessment of the quality of physics understanding expressed in the videos, we conclude that the authors of the fewer complete videos at the end of the semester do indeed demonstrate a greater physics understanding than both the authors of incomplete videos at the end of the semester and authors overall at the beginning of the semester.

This trend warrants explanation, though the work covered in this thesis only allows us to offer putative explanations rather than definitive ones. Figure 6.2 shows that the decline in the number of videos containing complete explanations is coincident with a decline in the number of videos offering any explanations at all (this is why the reported N-values are different pre and post). What happened to the missing explanations? We investigate this trend by examining the relative proportion of correct complete explanations (a.*.4) and flawed complete explanations (a.*.3) at the beginning and end of the semester, as shown in Figure 6.5.

We find that the incidence of flawed explanations among the complete explanations decreased from 68.7% to 36.4% over the course of semester, and that this decline is sta-

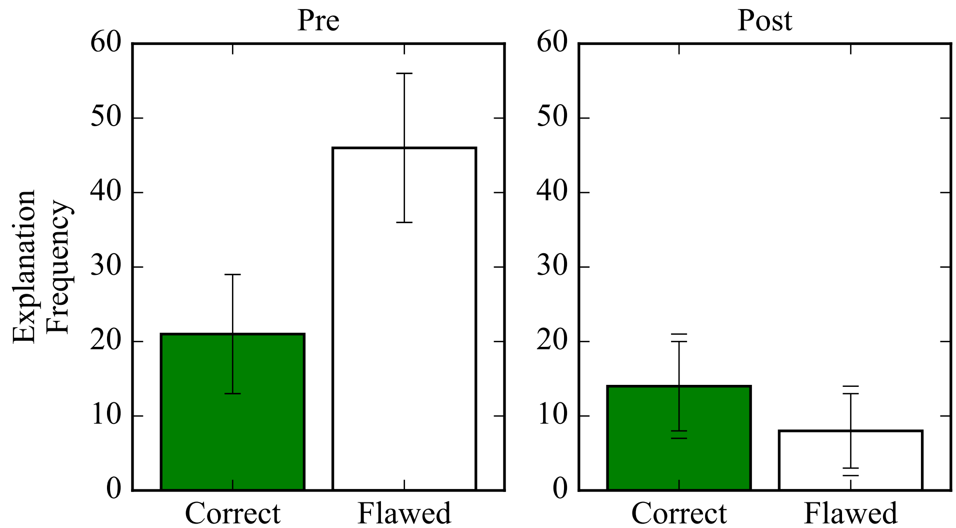


Figure 6.5: Frequency of correct, complete explanations (codes a.*.3 & a.*.4) and flawed, complete explanations (codes a.*.1 & a.*.2) in student videos at the beginning (pre) and end (post) of the semester. The number of complete explanations experiences an overall decline from pre to post (compare Figure 6.2). The proportionate decline in flawed explanations among the complete explanations from 68.7% to 36.4% over the same time period is statistically significant (Chi-Squared: $p = 0.015$).

tistically significant (Chi-Squared: $p = 0.015$) and coincident with the overall decline in complete explanations shown in Figure 6.2. This finding demonstrates that students in our research sample who offered complete physics explanations in their videos were relatively more likely to offer *correct* complete physics explanations at the end of the semester than at the beginning. One possibility admitted (but not demonstrated) by this finding is that some students struggling with physics concepts may have begun the semester by offering well-elaborated but flawed explanations of those concepts, and ended the semester by avoiding explaining those concepts altogether.

This change could be affected by a metacognitive development (students becoming more aware of the gaps in their knowledge) or by a rising awareness that flawed explanations yield lower ratings (see Figure 6.3); future work will be needed to explore this trend further.

6.4 Summary

When students' lab report videos were separated according to the completeness of the physics explanations contained within the videos' prediction vs. observation discussion, we found a significant pre/post decline in "excellent" ratings given by students to physics-content rubric items for videos with incomplete explanations (similar to our Chapter 4 findings), but not for videos with complete explanations. No significant pre/post decline in "excellent" ratings was found among student ratings of the non-physics-content rubric items.

Likewise, the typical student rating for physics-content items among the incomplete videos declined significantly, but did not change significantly among the complete videos ($p = 0.082$) from pre to post.

Expert assessments of those same videos showed a different trend; recall from Chapter 4 that overall expert ratings distributions did not change significantly from pre to post, and that we took this as evidence that students were, broadly, able to keep up with instructor expectations as the later labs became more difficult. In this analysis, no significant change in expert ratings distributions was likewise found among the incomplete videos, but among the *complete* videos, there was a statistically significant upward shift in the typical expert rating.

The student results suggest that, over the course of the semester, students became more attentive to the incompleteness of physics explanations in the prediction vs. observation discussion and/or learned to give explanation incompleteness more weight when assigning ratings. The expert results suggest that the completeness of prediction vs. observation explanations is associated more strongly with expressed physics conceptual understanding at the end of the semester than at the beginning. Our current analysis therefore admits the possibility that improved expressed physics conceptual understanding is a significant driver in the overall increase in student-expert rating agreement, though it cannot prove so definitively.

Students were also found to give relatively fewer incomplete explanations at the end of the semester than at the beginning, which suggests a change in students' attitudes toward which sorts of explanations merit inclusion in the videos they create.

In Chapter 7, we summarize these findings along with our findings from Chapters 3, 4, and 5, and present in some detail a future research agenda building off of our existing work.

CHAPTER 7
SUMMARY & FUTURE WORK

The previous chapters of this thesis described quantitative research into the rating behavior students demonstrated when engaging in peer assessment over time, and qualitative research into their retrospective accounts of their own cognitive and affective development during that engagement. We demonstrated a connection between the physics content of students' lab report videos and changes in student ratings of those same videos, and offered an explanation of those changes in terms of some of the themes uncovered during student interviews. This work constitutes our current progress toward developing a model for student engagement with peer assessment, and how this engagement changes over time.

In this chapter, we summarize our findings and describe how they could constitute the basis for a future research agenda for developing a model of student engagement with peer assessment.

7.1 Summary

In this section, we summarize our findings from previous chapters.

7.1.1 Background, Motivation, Design, and Validity of our Peer Assessment System

We developed a peer assessment system with algorithmic calibration [74, 50] and implemented it in an introductory mechanics course, offered both online and in-classroom over several years to several hundred Georgia Tech students. Students in this course all performed a series of laboratory exercises featuring computation and real-world observation-gathering, and each student produced a five-minute symposium-style lab report video for every lab. Students submitted these lab report videos to peer assessment (see Section 2.7), then assessed a set of practice and calibration videos before proceeding to assess each other's videos for a final grade using a rubric which we also developed (see Appendix A). We described our motivation for creating such a peer assessment system in detail in Chapter 1, we covered existing literature, our classroom setting, and our peer assessment system design in Chapter 2, and we covered our calibration algorithm, validity testing, and initial results in Chapter 3.

We determined that the grades produced by our peer assessment system were adequately accurate, but rather on the low end of the accuracy scale exhibited by other types of peer assessment. We attributed this result to the novelty and difficulty of the lab report video process, and the lack of student involvement in the development of the rubric.

7.1.2 Student Rating Changes Over Time

In Chapter 4, we described a quantitative improvement in student/expert agreement in lab report video assessments that persisted across several different offerings of the course. Student/expert agreement improved from the beginning of the semester to the end across both the set of training videos assessed by all students in common, and among a large, representative sample of the actual peer videos assessed by peers for a grade. Expert ratings of this latter set of videos remained relatively constant over the course of the semester, with experts most commonly assigning "fair" and "good" on a five-point poor-to-excellent rating scale.

Student ratings, however, changed substantially over this same period of time. At the beginning of the semester, students tended to give ratings on the higher end of the rating scale; at the end, students tended to give ratings in the center of the rating scale, with a substantial reduction in "excellent" ratings but no corresponding rise in "poor" ratings. A similar trend in rating behavior was found among students' ratings of training videos assessed by all students in common across three different course offerings, suggesting that students' engagement with our peer assessment system was broadly comparable across time and in different classroom settings.

This broad trend among students of increased agreement with expert ratings expressed itself differentially among the ratings assigned by students to different items on the rubric. In general, student ratings distributions came to resemble expert ratings distributions more closely among the two rubric items that dealt with rhetorical structure and production quality than among the three rubric items that dealt with physics concepts directly. Both parts of the rubric, however, did see a net improvement in student/expert agreement over the course of the semester.

This improvement in student/expert rating agreement was coincident with an improvement in students' mechanics conceptual understanding as measured by the Force and Motion Conceptual Evaluation (FMCE) [69]. This improvement was substantial, and was not significantly different from the FMCE improvement in our control course using the *Matter & Interactions* curriculum [13]. The coincidence of an improvement in student/expert agreement in lab report assessments and an improvement in FMCE scores equivalent to that of our control course suggests that our devotion of time and resources to peer assessment did not come at the cost of students' mechanics conceptual development.

We found similar trends in student rating behavior in different offerings of the course; the consistency of the change in students' rating behavior across semesters indicates that the conceptual and attitudinal changes underlying the behavioral changes are likely to be consistent across semesters, too.

7.1.3 Student Attitudinal Changes Over Time

In Chapter 5, we described our findings from clinical interviews with students who had participated in our course. These interviews were intended to collect students' retrospective accounts of their experiences with the course in general and the peer assessment process in particular, in order to identify themes in those accounts that might be corroborated with quantitative measures or amenable to further qualitative research

Many similarities were uncovered between different students' accounts, but two themes in particular seemed to be potentially relevant to the observable changes in student rating behavior. Interviewed students consistently reported feeling an inclination to give "nice" or high ratings to each other at the beginning of the semester, and reported developing a more critical attitude toward the rating scale and/or a greater attention to detail and critical disposition toward physics errors as time went on. The vast majority of interviewed students (eight out of ten) reported becoming more willing to give "lower" ratings and/or feeling more critical at the end of the semester—but only one student reported becoming more comfortable with giving "poor" ratings specifically. As shown in Chapter 4, these self-reported attitudinal and/or behavioral changes were not generally coincident with an increase in poor ratings among students.

The second theme consistent among the interviewed students' accounts is one relating to the role of expertise in lab report video assessment. Students reported feeling inexpert relative to the course instructors in physics conceptual understanding and in experience with video assessment, and students claimed that the instructors' tendency to give lower ratings than they did was related to the instructors' relative expertise in these matters. Students frequently mentioned they believed instructors to have a greater ability to attend to small mistakes in the videos, more familiarity with the rubric, and/or a stricter assessment of the qualifications for the "very good" and "excellent" ratings. Some students reported they were able to develop what they considered a more expertlike attitude toward the rubric and rating scale over time, thereby improving their agreement with the expert assessments of the practice and calibration videos. Other students expressed no such improvement, or stated that the expert ratings seemed arbitrary and/or unpredictable.

Some interviewed students also expressed a particular practice of beginning their peer assessments by mentally assigning a high score to the video when they started watching, then mentally subtracting points every time they noticed a significant error.

7.1.4 Student Communication Changes Over Time

In Chapter 6, we described our analysis of the communicative content of students' lab report videos, and the way this content influenced the ratings which students gave over time. We focused our study on a particular section of their lab videos where students discussed a comparison between an observed and a predicted trajectory, and we examined these discussions at the beginning and end of the semester. We classified students' lab report videos based on the length & sophistication (the "completeness") of the physics arguments they presented when comparing their predicted trajectories to their observed ones, and explored how peer and expert ratings of those videos changed with completeness over time.

We found that student ratings of videos with complete explanations improved significantly over time, suggesting that the completeness of explanations became a salient factor to students' rating decisions as time went on. We also found that students tended to offer fewer flawed explanations of physics concepts in their videos toward the end of the semester without offering more correct explanations, suggesting perhaps that students learned to avoid offering any explanation at all for concepts which were unclear to them—though more research is needed to substantiate this trend.

7.2 Future Work

In this section, we enumerate components of a possible research agenda for expanding upon the findings presented in this thesis and summarized above. We describe the limitations of our current inquiry, the corpus of data which that inquiry produced, and new analyses that could use that data to push our lines of inquiry even further. We also offer several new, exploratory lines of inquiry unaddressed by our existing work that may be initiated using our gathered data. Subsections are arranged in order of their presentation in earlier chapters of this thesis.

7.2.1 Validity of Our Peer Assessment System

In Chapter 3, we established the validity of our peer assessment system as compared to other systems, and identified several factors in the literature known to contribute to higher peer assessment validity. Here, we offer two possible ways to improve validity, and an associated short-term research agenda.

Classroom Practices

We have run our peer assessment system with a single version of SWAPR, the same rubric, the same training videos, and the same general classroom practices (TA regrades, face-to-face feedback with different people than the final assessors, etc.) for several semesters. This particular configuration of peer assessment factors is therefore well-documented by our research, and further insights into student engagement with this system could be expected from analysis of our existing data. On the other hand, we could expand the breadth of our peer assessment research and possibly improve our peer/expert agreement by incorporating more of the salutary factors identified by Falchikov & Goldfinch [19] listed in Section 3.2 into our classroom practices. Would including students in rubric development improve buy-in and/or accuracy? The use of a single, global assessment based on overall criteria rather than a categorized multi-item assessment is also indicated as beneficial by the list in Section 3.2; if we were to compare the results of a single-item assessment to those of our existing multi-item rubric, would the single-item assessment produce better accuracy? If so, would it also be an effective means for superior feedback?

These research questions could be addressed in future offerings of the course with a new round of rubric design, and could achieve results in a relatively short time.

SWAPR Development

Our calibrated peer assessment system SWAPR (described in Chapter 2) could stand to be improved, independent of classroom practices or setting. Crucially, no attempt is currently made to optimize the assignment of students' lab report videos to their peers; assignment in SWAPR is completely random. A more sophisticated assignment procedure could, for instance, aim to ensure that every video is assessed by a group of peers of roughly equal average calibration score. This would, presumably, help to reduce whatever variability in final grade accuracy is caused by variability in peer group quality (as measured by calibration score).

Developing an assignment algorithm which optimizes for peer group quality—either by ensuring a uniformity of group calibration score across the whole section, or balancing some other performance metric or demographic feature—is an obvious next step in the further development of SWAPR. The construction of such an optimized group of agents belongs to the class of *stable matching problems* in computer science [75], which are well-explored in existing literature.

7.2.2 Student Rating Changes Over Time

In Chapter 4, we discussed a quantitative trend toward greater peer/expert agreement by describing changes over time in peer ratings distributions compared with expert ratings distributions. The scope of this analysis, however, was limited just to student and expert ratings of peer videos and training videos, and was concerned with establishing trends in peer/expert agreement *per se*. By instead recasting peer/expert agreement as a dependent variable, we can open our inquiry to explore the effects of the many other quantitatively-measured factors in our classroom relevant to peer/expert agreement, and how these effects change over time.

Our students produced far more material than that which was directly related to the peer assessment assignments; they took exams, answered clicker questions, filled out Course Instructor Opinion Surveys, and so on. Each student also had a full registrar file, with demographic information, incoming and outgoing grade point averages, choice of major, course of study, etc. Some of these factors stayed constant over the semester, others changed with time.

To outline a full model of student engagement, including change over time, it may be helpful to perform a stepwise regression analysis of our collected student data to find the factors most predictive of student/expert rating agreement, given our finding that student/expert rating agreement changed substantially during instruction and was associated with changes over time in student communicative behavior. Dividing this analysis into pre and post regressions would also allow us to examine the change over time in the predictive power of these factors.

These analyses could be performed (after some careful design work) with data already on-hand (105,395 ratings and comments given by students to these peer videos), and constitute attractive targets for continuing our research in the near term.

7.2.3 Student Attitudinal Changes Over Time

In Chapter 5, we described the themes and trends expressed by students in their retrospective accounts of their experiences with our peer assessment system—attitudinal and otherwise. Our qualitative investigation into the experience of peer assessment used only recorded interviews performed by ten students recounting their past experiences with peer assessment in a clinical setting. This work yielded interesting findings, but suffered from the artificiality of the environment in which it was conducted. As yet, we have no direct insight into the subjective experiences of students participating in our peer assessment system under naturalistic conditions besides these retrospective accounts. A more thorough investigation into the qualitative experience of engaging in peer assessment would need to examine peer assessment when it is actually practiced in the classroom.

A live study of students participating in actual peer assessment would allow us to assess the accuracy of these interviewees' accounts, and connect student rating behavior with the lived experience of engaging with our peer assessment system.

As regards existing data, further inquiry into students' self-efficacy and their perception of themselves as physics students could proceed from a deep reading of other parts of the interview transcripts (Appendix F), where students describe their relationships to other students, their expectations for the course, and other topics particular to each interviewee. Our data on-hand includes 15 recorded and transcribed clinical interviews with students, covering physics reasoning, think-aloud video assessment, and retrospective accounts of their engagement with peer assessment, totaling $\sim$15 hours.

7.2.4 Student Communication Changes Over Time

In Chapter 6, we described changes over time in student communicative practices as expressed in their lab report videos, and how those practices correlated with changes over time in student rating behavior. Here, we lay out two unfinished lines of scientific-communication-related inquiry which could be pursued with existing data (1799 lab report videos created by 388 students, totaling $\sim$150 hours; 301 videos are coded and partially transcribed).

Lab Report Video Coding

Our current research used only a portion of the content of the first and last lab report videos (the "prediction vs. observation discussion", where students compared their simulated results to their observations and explained any discrepancies between the two). Other portions of the lab report videos should be amenable to study with a similar coding scheme, and can be expected to yield insights into our students' communicative practices, the way they construct computational models, their understanding of fundamental physics principles, and the way these activities interact.

More immediately, only a few of the codes developed for our exploratory research (see Appendices C & D) showed significant results in our research as described in Chapter 6, but there is every reason to expect that a different investigation of the other codes may yield interesting correlates and important trends over time in the way students communicate. These codes—having already been assigned—represent probably the most obvious next target for future research into student communication practices, and how they change over time.

Linguistic Analysis

Our work has produced a large corpus of interview transcripts and recorded speech. Some initial insights into the development of students' use of scientific language during our course should be available through straightforward lexical analysis; our own first attempt at investigating language usage *per se* in students' lab report videos (examining rating changes by gender among videos using "hedged" language—see code a.E in Table D.1) yielded no significant results, but other analyses (*e.g.*, the use of nominalization [76]) might produce interesting findings.

7.2.5 New Exploratory Research

Here, we describe two possible avenues of research that could be conducted starting with existing data, but which have not been explored during the undertaking of this thesis.

Gender Analysis

A study of gender role performance in our unsigned but not anonymous peer assessment system should yield some novel insights into the construction and performance of gender roles in science classrooms generally. In particular, the null result with respect to gendered usage of hedged language described above in Section 7.2.4 may itself be worthy of further investigation, given previous literature finding women's speech, particularly in science-classroom settings, characteristically less confident and more noncommittal than men's speech in the same settings [77, 78].

Clickstream Analysis

During the course of our research, we gathered a dataset of student interactions with the online video player they used to watch their peers' lab report videos ("clickstream" data), allowing us to reconstruct their engagement with the video; what parts of the videos they played, whether they re-watched certain sections, frequency and location of pauses, etc. Our group has previously described a method for reconstructing students' video engagement from their clickstream data [79]. This data exists for students' peer evaluations in 2016 and 2017, and for students' training phases in all semesters, and has been collated but not thoroughly investigated.

Are there certain viewing behaviors associated with higher peer/expert agreement? Are certain visual or audible features in students' videos more likely to draw frequent attention from peer viewers? Our clickstream data constitutes a promising basis for formulating research questions about student engagement with peer assessment not addressed in this thesis.

7.3 Final Remarks

This thesis began with research questions about the efficacy and accuracy of peer assessment, and the ways in which students engaging in our peer assessment system learned to produce and critique scientific communication. Our work has demonstrated to our satisfaction that our peer assessment system (which includes a reasonable grade-dispute process) can and does produce grades accurate enough to replace instructor grading, that students do learn to become more accurate raters given more experience with our peer assessment system, and that this improvement in accuracy appears to be associated with increased student attention to to a feature of physics communication (*i.e.*, explanation completeness) that instructors consider important and salutary.

This work did not arrive at a complete synthesis of the factors involved in peer assessment, nor did it elaborate a predictive model of student engagement with peer assessment. Rather, we intend that this work will serve as a validation of the use of our peer assessment system in an introductory mechanics course, in terms of both logistical utility and the expansion of introductory learning goals to include scientific communication and critique; and we intend that this work will provide a novel insight into the development of student peer assessment practices and attitudes over time through the demonstration of broad, statistically significant quantitative changes in student ratings. Finally, we intend the data collected and the findings described during this undertaking to serve as the basis of a future research agenda for more fully understanding student engagement with peer assessment in the physics classroom.

Appendices

APPENDIX A

RUBRIC

The full rubric used in this study for peer assessments, as it appeared to students.

Criterion	Marginal		Advanced
Organization Structure	- No clear organization/logical structure - Lacked introduction and/or conclusions - Introduction fails to present problem - Introduction lacks a statement of result - Introduction lacks preview of major sections - Few or no transitions or signposts - Viewer is "lost"	Poor | Fair | Good | Very Good | Excellent Comments:	- Excellent organization/logical structure - Excellent intro and conclusion - Introduction clearly states problem - Excellent statement of result in introduction - Introduction contains excellent preview of major sections - Excellent use of transitions and signposts - Speaker helps audience follow structure
Content Models	- Fails to identify models relevant to the physical system - Lacks discussion of main physics ideas - Lacks discussion of application of ideas - Lack of connections between fundamental physics principles and the model	Poor | Fair | Good | Very Good | Excellent Comments:	- Identifies models relevant to the physical system - Excellent discussion of main physics ideas - Excellent application of ideas to problem - Excellent connection between fundamental physics principles and the model
Content Prediction Discussion	- Data used to initialize the model are not clearly identified - Fails to discuss how parameters are adjusted to fit data - Fails to discuss how data does or does not fit the model - Fails to discuss whether the computational model does or does not predict the motion of the object observed	Poor | Fair | Good | Very Good | Excellent Comments:	- Data used to initialize the model are clearly identified - Excellent discussion of how parameters are adjusted to fit data - Excellent discussion of how data does or does not fit the model - Excellent discussion of whether the computational model does or does not predict the motion of the object observed
Content Overall	- Discussion contains major physics errors - Lack of discussion of "What if..." question - Lack of discussion of "What does it mean" question	Poor | Fair | Good | Very Good | Excellent Comments:	- Discussion contains no physics errors - Excellent discussion of "What if..." question - Excellent discussion of "What does it mean" question
Production Delivery	- Poor lighting/low resolution; video hard to see - Poor audio; audio difficult to hear - Distracting vide (e.g., shaking) - Overall production quality is poor - Visuals fail to enhance message - Poor vocal qualities in narration	Poor | Fair | Good | Very Good | Excellent Comments:	- Excellent lighting/resolution - Excellent audio quality - No distracting video - Overall production quality is excellent - Visuals reinforce and support message - Excellent vocal qualities in narration

APPENDIX B
HOW PEER GRADING WORKS

This four-page document was given to students starting in Spring 2014 to explain the calibration process for the peer assessment system.

Why do we do peer grading?

- We do peer grading because peer review is a very important part of real science, and we want this course to help you learn real scientific practices! That's also why we teach computational modeling, and why we do video lab reports; real physicists do a lot of computation, and all scientists have to be good at reporting their findings to their peers.

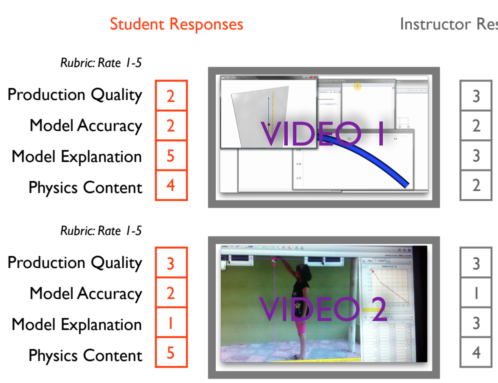
What is calibration, and why do we do it?

- Some people have learned to be better at evaluating videos than others. We don't want your grade to suffer if your video happens to be assigned to an inexpert evaluator, so we "calibrate" everyone with a calibration assignment. People who score well on the calibration assignment have more influence on the videos they grade, and people who score poorly have less.

How You Get Your Calibration Grade

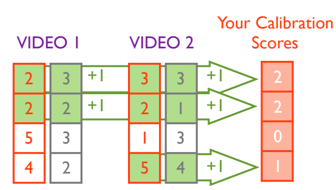
1. Let's say your calibration assignment consists of two videos, and you evaluate them according to a four-item rubric.

2. Earlier, the course instructors collectively evaluated these same two videos.



(Everyone else evaluates these videos, too.)

3. We match up your evaluations to the instructor evaluations. When your response for an item is within one point of the instructors' response, you "agree" with the instructors, and you get a calibration point for that item. Congratulations!



In this example, you agreed with the instructors on item #1 for both videos, so your calibration score for item #1 is 2. Good for you! Likewise, you got 2 calibration points on item #2. On item #3, you did not agree with the instructors either time, so you got 0 points; this suggests you might have some trouble judging item #3, the model explanation. On item #4, you agreed on one video but disagreed on the other, so you only got 1 point.

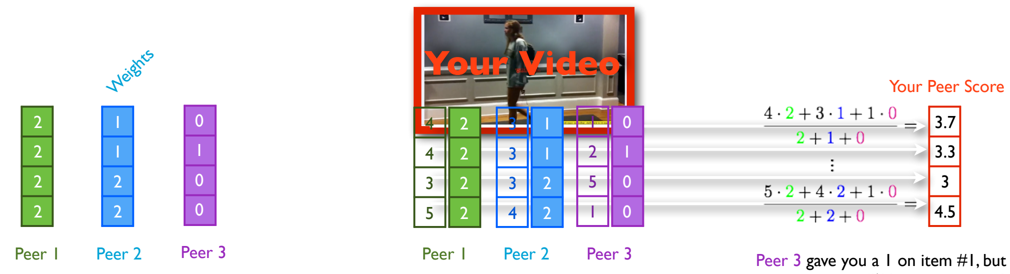
Your calibration grade is based on how many points you earned: in this example, you earned 5 out of 8 possible points, so your calibration grade is 5/8 = 62.5%

How You Get Your Peer Evaluation Grade

1. Let's say your video got sent out to three of your peers. Each of these peers was calibrated, and has a set of calibration scores shown in solid boxes.

2. Now, each peer fills out the rubric for your video. Their responses are shown in empty boxes.

3. For your final peer grade on your video, we take the weighted sum of these 3 responses, as below.



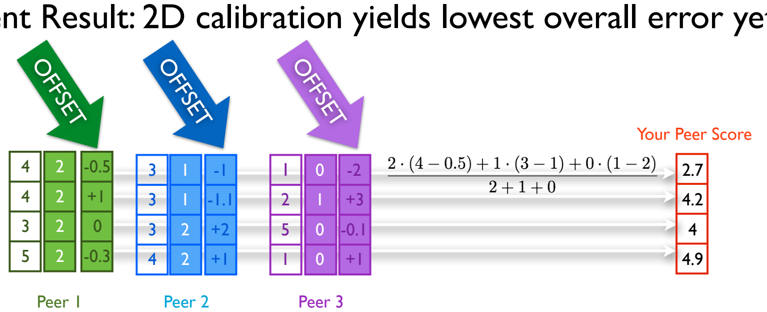
Peer 1 is a solid evaluator all-around; this person agreed with the instructors on every single item for every single video.
Peer 2 was good at judging the last two items, but not quite so good at the first two.
Peer 3 is a poor grader. Peer 3's response will have very little effect on your grade.

Note how Peer 3's scores are very different from Peer 1 or Peer 2. You might not have gotten a fair grade from Peer 3.

Peer 3 gave you a 1 on item #1, but that didn't affect your score at all! Your score was dominated by the contributions of Peer 1 and, to a slightly lesser degree, Peer 2, because they had better calibration scores.
The final peer grade you see on Webassign is just the percentage of possible points you earned: in this case, you got 14.5 out of 20 possible points, so you got a 14.5/20=72.5%

"Offset"-Style Calibration

Recent Result: 2D calibration yields lowest overall error yet



Calibration assigns each student both a weight *and* an offset for each item (offset = average difference between student response and expert response). During peer grading, this offset is subtracted from each student response before averaging.

APPENDIX C

COMPARISON CODES

We found that students, when comparing their prediction and observation plots, would sometimes compare them by describing the *similarities* between the plots (*e.g.,* “The model fits the actual results very well.”, “In both the observation and the prediction we could see the change in velocity was the same...”), sometimes by describing the *differences* between the plots (*e.g.,* “...throughout time, it begins to deviate more and more.”, “It was shown that there was a clear discrepancy between the observed and predicted positions.”), and at other times by describing both. We therefore constructed the codes in the “Agreement” and “Discrepancies” columns of this table to accommodate these two different sorts of explanation.

The “Specifics” column encodes certain properties of the relationship between observed and predicted motion that appeared regularly among student descriptions, but did not solely have to do with the magnitude of agreement or disagreement between them.

Table C.1: Comparison Codes. Each code classifies the *author's* description of the comparison between predicted and observed trajectories.

Agreement (describing the similarity)	Discrepancies (describing the difference)	Specifics
c.1.1 The observed and predicted motion agree almost completely. Includes a superlative phrase like “spot on”, “pretty much perfect”, or “nearly identical”, or flatly states “they are the same”	c.2.1 There are almost no discrepancies. Includes an infinitesimal phrase like “almost negligible”	c.3.1 The model overpredicts or overestimates the observed data, or is otherwise “more than” the observations in a specified way
c.1.2 The agreement is very good. Includes an emphatic phrase like “really good” or “very close”	c.2.2 The discrepancies are very small. Includes a diminutive phrase like “slight differences”, “small”, “a little different”, “	c.3.2 The model underpredicts or underestimates the observed data, or is otherwise “less than” the observations in a specified way
c.1.3 The agreement is moderate. Includes a moderate phrase like “pretty similar”, “okay”, “relatively good”, “not exact”, or “for the most part”, or “not quite a match”	c.2.3 The discrepancies are moderately large. Includes a moderate phrase like “relatively large”	c.3.3 The observed and predicted motion diverge as time proceeds
c.1.4 The agreement is poor. Includes a negative phrase like “not very good” or “quite off” , or a flat “they did not match”	c.2.4 The discrepancies are large. Includes an emphatic phrase like “big difference”, “obvious” or “very large”	c.3.4 The observed and/or predicted motion are described as having specific properties, such as being “linear” or “periodic”
	c.2.5 Discrepancies are present but their magnitude is not described.	

APPENDIX D

ATTRIBUTION CODES

We found that students, having described the comparison between their prediction and observation plots, would then tend to attribute the results of their comparison (usually a discrepancy between the plots) to a limited number of physics factors. We list these factors in the "factor" column.

Students would then offer an explanation for how that factor produced their stated result; these explanations would vary in substance and quality. We developed a heirarchy for the completeness and correctness of these explanations. "Explanation" codes 1 and 2 are incomplete because they lack a clear statement of cause and effect, and are of indeterminate correctness because whatever physical understanding intended by the student is only imputed, not explicitly stated. Explanation codes 3 and 4 are both complete, because they contain a clear statement of cause and effect; they differ in correctness, as 3 contains vague or flawed physics and 4 containst clearly stated, correct physics. The "Explanation" column of codes, unlike the other codes in this table or in Appendix C, encodes *evaluator judgment* about students' speech and not objective features of the speech itself.

Table D.1: Attribution Codes. The factor codes classify the factors to which the *author* attributes their prediction vs. observation results. The explanation codes classify the quality and completeness of the explanation of each of these factors, in the judgment of the *coder*.

Factor	Explanation	Other
a.1 Friction	a.*.1 Listed only	a.A Newton's Second Law is proven correct/accurate
a.2. Drag (from air)	a.*.2 Factor is expounded upon, but incompletely. Link between factor and effect on system is vague, or effect is non-specific, such as a "discrepancy". Source of factor is identified, and/or action which the factor does.	a.B The model is proven correct/good/accurate
a.3 Human Error in Tracker or experiment	a.*.3 Factor is linked to a specific effect with vague or misconceived physics/reasoning	a.C Early errors persist or compound with time/the model is sensitive to initial conditions or calculated parameters
a.4 Initial Force	a.*.4 Factor is linked to a specific effect with correct physics/reasoning	a.D Model is "ideal/perfect/not the real world", and/or observations are in "real world"
a.5 Gravity		a.E Uses hedged language like "probably", "maybe", "could be", "perhaps", "if there were", "I think" (but not "mainly" or "mostly")
a.6 Normal force		a.F Mentions a way that the model could be made more accurate
a.7 Shape/texture of object/surface		
a.8 Path of object is unusual/imperfect/ multidimensional		
a.9 Framerate/Recording error		
a.10 Other/unknown forces or factors (not specified)		
a.11 Energy transfer		
a.12 Rounding/numerical error		

APPENDIX E

PRE/POST INTERVIEW SUMMARIES

Here, we summarize each 2017 interviewee's statements relevant to the research questions above and include relevant quotes. For full transcripts of all interviews, see FIXME APPENDIX B. Students A, B, and E were white and female. Student C was an East Asian male, and Student D was a white male.

E.1 Student A

In the pre interview, Student A specified that she tended to give the authors of the videos she reviewed the "benefit of the doubt", and tended to "focus on the positives" before considering the negatives. She attributed this attitude to a sense of self-worth; she liked "to have a positive mindset about others' work, because [she also has] a positive mindset about [her] work." She anticipated that her tendency to give people the benefit of the doubt would "correct itself" once she assessed more videos.

In the post interview, Student A identified a document available on Piazza [80] (the online question-and-answer forum used in our course) as very helpful in distinguishing between the criteria for adjacent ratings (e.g. what merits a fair vs. a good). She reported this document as helping her get better calibration grades, since it had previously been straightforward to tell when the video is excellent or poor, but more difficult to assess videos of intermediate quality. No other interviewee mentioned this document, which was an early draft of the rubric with a description of each rating for each item.

E.2 Student B

In the pre interview, Student B claimed she was a "really easy grader" and "just not a very critical person" who "hates" giving ratings below good because it "makes [her] feel bad". She felt people should get a good for at least trying; to justify a poor rating, she would need to be able to "blatantly tell [the author] just did not care at all." Conversely, when assessing videos, she was "always trying to find how [she] can give a better grade". Like Student A, she expected that she may become harsher over the course of the semester. She had the impression that her ratings for the training videos usually ended up one rating higher than the instructor ratings.

In the post interview, Student B still reports that grading someone negatively still "feels bad", related to the fact that she "doesn't have a hundred percent qualifications" and is not a physics professor. She said her reluctance to give poor grades didn't change, but she doled them out more frequently because she "understood the rubric more". She specified that the high-stakes nature of the peer assessment exercises "put a lot of pressure" on her and was stressful; if the nature of the peer assessment had been low-stakes formative feedback, she "wouldn't [have had] any trouble."

E.3 Student C

In the pre interview, Student C identified the instructors as probably having "a harsher grading system [than students], but [he doesn't] think that's necessarily a bad thing". With respect to assigning ratings, he believed in "innocent until proven guilty" and mentally started each video off with a grade of 100, subtracting points every time he saw a "glaring" or "eyecatching" error. This simply felt to him like the right way to grade, and let the recipient of the grade leave the assignment on a "more positive note". Less dramatic errors are problematic; he finds it difficult to "pinpoint all the errors, but... very easy to pinpoint some of the errors." He believed instructors had "a more thorough understanding of the material", which enabled them "to analyze a lot more deeper into questions" than students. His own ratings on the training videos would often only be one apart from the expert ratings, and he thinks students might have been inclined to give higher ratings than experts because of the rubric's ambiguity regarding the central ratings on the rating scale.

In the post interview, Student C reported he improved his calibration grades (i.e. his agreement with instructor ratings on the training videos) after eventually realizing that each video is more or less structured like the rubric itself, in sequential sections. After recognizing this and adjusting his rating strategy, his ratings became much more instructor-like, so he believed this must have been the rating strategy instructors themselves used. He also reported having developed "more knowledge into grading the scale" based simply on having rated many videos. The very good and very bad videos have a clear pattern of what they're doing right and wrong, he said. He also said that physics content knowledge is most important to the Content: Models rubric item (#2) because "without knowing physics content, it's hard to do anything about that". He did not report becoming a harsher rater, *per se.*, simply that his ratings became less "inconsistent" with the instructor ratings after he alighted upon the strategy of mentally dividing each video into sequential sections corresponding to the rubric items.

E.4 Student D

In the pre interview, Student D expressed frustration that the experts' comments on the training videos seem to indicate a higher rating than they actually assign; he didn't always understand why the experts "dock as many points as they do". He himself leaned "toward giving someone credit for something, rather than not giving them credit", and believed that "credit should be given when a student shows that they know the concept". He specifically phrased some of his thoughts about rating harshness in terms of respecting the text of the rubric, and not requiring students to "exceed any expectiations [he had] for them to earn the points they deserve"—again repeating frustration with the experts giving lower ratings than their comments would indicate. In his earliest video assessments, he graded "leniently" and gave higher ratings than the experts, who were "harsher" than him and whose "definition of fair and good" clearly differed from his. After this experience, he "swung more to the slightly harsher" side of things. His attitude toward giving or deducting points was "[his peers] definitely have to earn their points, but [he's] trying to see where they're succeeding."

In the post interview, Student D attributed his early lenient attitude to "mainly sympathy towards my fellow student." He had thought that if he "were in the position of the person being graded, [he'd] want the grader to give [him] credit wherever possible." At the time of the post interview, he still sympathized with his fellow students, but after gaining experience with the peer assessment system, he spent more time thinking about how he could "most accurately match the system" and less about "how [he's] supposed to grade [the videos]". The full context of this statement implies he was thinking less about what sort of grades he felt his fellow students deserved, amd more about how he would maximize his accuracy with the expert ratings:

> Interviewer: Why did you think you began the semester with that attitude of trying to find reasons to give people credit? Where did that attitutde come from?
>
> Student D: Just mainly sympathy towards my fellow student. Like, I thought that if I were in the situation of the person being graded, I'd want the grader to give me credit whenever possible, whatever is possibly due. And I was just trying to be a reasonably fair but not ridiculously easy grader.
>
> Interviewer: And how would you describe your feelings of sympathy now, at the end of the semester? Or rather, you know, back around lab four—around the end of the semester.
>
> Student D: So, I still find myself... once I found out how the grading was figured out, where your decisions are weighted based on how well you matched

> the instructors' guidelines in the calibration activity... I mean internally, I still felt like "Oh, these are my fellow students", but I more thought of how I'm going to most accurately match the system, as opposed to how I'm supposed to grade these.
>
> Interviewer: So, you're interested in matching the system. You're interested in anticipating what the instructor responses are going to be. What do you think motivates the instructors to set the responses they did?
>
> Student D: For me, it's kind of hard for me to really know, considering... I dunno, I don't really have much grading experience. I guess that the biggest ideal instructors are trying to achieve is objectivity. For me I think my problem was there was a lot of gray areas, to me, and so I was more lenient at the start. And for them, it's... "Is it right, or is it wrong?"

In the post interview, he described the instructor grading as "a bit unfair and super strict". Toward the end of the semester, he moved away from "looking for reasons to give [authors] credit" and toward requiring "repeated, consistent achieving of the criteria".

E.5 Student E

In the pre interview, Student E described the instructors as being "inconsistent with what they rate harshly, and what they don't rate harshly". She said the instructor evaluations were nothing like hers, and she was clearly frustrated that she found it very difficult to match the instructor ratings and "[didn't] really know what to do about that." (She was, however, pleasantly surprised that she matched closely with the instructors in her think-aloud video evaluation.)

She attributed this inconsistency in instructor ratings to the lack of an individual rubric for each lab, specific to the content of that lab. Without such a rubric, "we don't know what we're looking for... how are we supposed to judge other people's labs?" For this reason, and because students are asked to begin making their next lab video before they've received feedback from their previous one, she did not think that the peer evaluation system as implemented in her course was a reliable system for assigning grades.

In the post interview, Student E expressed the feeling that her calibration grades did not significantly change over the term. She said she developed a better understanding of how the lab report videos tend to be organized, but that this has only helped her do her evaluations more quickly, not more accurately. When evaluating the same video as in the pre interview, she again expressed surprise at how similar her ratings were to the experts, but pointed out that for Item #4 (Content: Overall), she and the experts "had the exact same reasons [stated in the comments]" but gave different ratings, which illustrates how the peer assessment system is "not good". These are the ratings and comments in question:

- Instructor Rating: *Good* - Inadequate WDIM[1] explanation—no motion/net force connection. Good WI[2]. Physics principles appear to be soundly understood, but not always clearly expressed.

- Student Rating: *Fair* - addressed what if question on axes, however described the axes as 'straight' which is irrelevant because an axis is already a line. He should be describing further direction and why he chose his starting point. → he said that it moves in the positive x direction. - He mentions that if the axes were flipped, as the question asks, then the velocity would not change in magnitude but only direction - He addressed the second what if question [3], and answered correctly that the individual forces cannot be determined, however, his explanation was incorrect ("invisible and tough to determine") when the correct answer would be to explain forces such as

[1]The conceptual "What Does It Mean?" (WDIM) question that students were instructed to answer in their lab report videos for this lab (Lab 1) was "Your observations can be predicted well by your computational model when you set the net force equal to zero in your model. Is it possible for you to say how many pushes and pulls are added together to give zero net force in the case that you observed? Discuss briefly."

[2]The hypothetical "What If...?" question for Lab 1 was "When you analyzed your video data in Tracker, you made a choice of the orientation of the axes. Discuss briefly how your results would change (and how they would remain the same) if you were to flip the axes."

[3]Actually the "What Does It Mean?" question.

friction, weight force, etc. that we didn't measure.

Student E said specifically that she gave a "fair" because "he got the first question right but not the second." She also noted that the unreliability of instructor ratings was accompanied by unreliability among peers who left similar comments but different ratings for her videos—likely for the same underlying causes, relating to lack of specificity in the rubric.

Student E, unlike the other interviewees, spoke extensively about her experience requesting and receiving TA regrades for her lab videos. She believed students were *harsher than* TAs, because she would consistently get higher grades from the TAs than from her student peers. She distinguished the TA assessments from the "instructor" assessments of the training videos (which she said were simply "inconsistent" with her own, rather than tending one way or the other).

APPENDIX F

INTERVIEW TRANSCRIPTS

F.1 Student 1, Spring 2014

Interviewer [01]: Okay, excellent. So thank you so much for agreeing to participate in this interview. Before we begin, I've sent you a consent form I'd like to sign, along with a media release form. Just sign them electronically. The consent form tells you the nature of this study, and that there is no known risk to you for participating in this study. The media release form, if you choose to sign it, gives me permission to share anonymous recordings of this interview with other researchers for research purposes only. You don't need to sign the media release form to participate in this study, but you do need to sign the consent form if you choose to participate in this study.
Student 43]: Okay.
I [00:43]: We can go over that now... if you have any questions.
S [00:48]: Um... no, can I just send it after this, or...?
I [00:52]: Uh, actually I do need to collect consent before we begin the... before we begin the program.
S [01:00]: Okay, um...
I [01:03]: You should just be able to like, click on the form and just type in your name. That's all that's needed.
S [01:11]: I can't type it in. Uh, sorry, hold up, let me try it...
I [01:16]: Are you on a Mac?
S [01:18]: Yeah.
I [01:18]: Uh, yeah, you should be able to... like if you use the text tool in the top toolbar, you should just be able to type anywhere on the pdf that you want.
S [01:28]: Okay, I'm not...
S [01:33]: Um, I don't have the text tool on the top menu... the bar.
I [01:37]: Interesting. Do you have it open in Preview?
S [01:43]: Oh, nope, I have it open online. Ah. Let me open it in Preview.
S [02:18]: Alright, I'm sending it now.
S [03:05]: Okay, I just sent it!
I [03:08]: Okay. Excellent. So that was the consent form. The other thing is the media release form, which covers a different thing. The consent form is just informing you of the risks of this study, of which there are none, and the goals of this study, which is to understand what Ss are thinking when they do these peer evaluation exercises.
S [03:26]: Okay.
I [03:26]: The media release form, uh, says, that when I'm speaking at a colloquium or showing presentations to other researchers, I might from time to time want to show clips of videos from these interviews. It is totally within... it is your decision whether you [glitch]
I [03:58]: [unintelligible]
S [04:05]: Oh, you're freezing up, sorry.
I [04:06]: Uh, okay, I'm sorry, can you hear me now?
S [04:08]: Yeah. The video's still frozen.
I [04:14]: ...think [glitch] restart. Uh, my computer does seem to be freezing up. Ah, I wish this would run a little more smoothly. Uh, stay—please stay on the line and I will... I will join you shortly.
S [04:30]: Okay!
I [04:30]: Thanks for your patience.
[S sends and receives messages online while I reconnects]
I [05:59]: Okay, I'm back. [glitch] This should go a little more smoothly.
S [06:03]: Okay.
I [06:03]: Okay. The media release, you can in fact make that decision afterward, so you know what actually has gone into this video. Um, so, with that in mind, and now that I've received your consent form, uh, let's go ahead and get started.
I [06:23]: I'd like to begin by walking you through a think-aloud problem for practice. Then I'll ask you to evaluate a sample lab video for me while you think aloud, just like your lab evaluation exercises, so I can understand what you think about the video and why you're giving the ratings you do. Next, we'll discuss that experience and your experience in the course so far. Finally, I'll ask you to answer some follow-up questions. The whole interview we expect will take about an hour. I'll be recording this interview over your webcam, and when you're doing the evaluation, I'll ask to record your computer screen as you're watching the video. Would you like to begin?
S [07:01]: Yes.
I [07:02]: Okay, excellent. So the first question is not going to be related to the videos as such. Uh, I want to give you a warm-up on how, uh, think-aloud interviews work. Okay, so please get something to write with if you don't have... if you don't have it already.
S [07:19]: I have some.
I [07:19]: Okay. Uh, so I'm going to ask you a physics question, and I'd like you to narrate to me the the thoughts going through your head as you solve it. The thoughts that you speak to me don't have to be coherent. They don't even have to be in complete sentences. I'm not going to ask you any further questions during the process, and I will mostly remain silent. From time to time, if you're not talking for a while, I may prompt you to keep talking.
S [07:47]: Okay.
I [07:48]: Does that make sense?
S [07:49]: Yes.
I [07:49]: Would you like to begin?
S [07:51]: Yes.
I [07:52]: Okay. this is a two-part estimation question.
S [07:55]: Okay.
I [07:56]: What is the volume of the room you're in right now, and if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for that room to fill with water?
S [08:09]: Um, okay, so first... find the volume—I'm in like a square sort of room, so the volume of a square... uh, formula... [checks on computer] Uh... honestly, I'd look that up. Um, and...
S [08:32]: Volume of a cube? Sorry. And then it would be like, sides cubed, right? And then I guess measure the room? If I really have to. Uh, I have those little tape measuring thingies. I guess [gestures across room] measure this? The part behind me, the part over there, and the part over there, and then find the volume of the room, and um... what was the second part?
I [08:54]: The second part was if you were to turn on an ordinary kitchen faucet in that room, about how much time would it take for the room to fill with water?
S [09:05]: Okay. Uh... [shakes head] I guess that depends on the pressure of the kitchen faucet. Um... oh! wait! Or the rate of time the water's flowing. I guess taking the volume from the last question and then putting that as the desired volume that you'd want the water to fill. And, maybe like, turning the faucet on for a little bit of time and seeing how much water comes out, volume-wise? And using that proportion to see how much the room fills up.
I [09:41]: Okay. Very good! That's your answer?
S [09:43]: [shrugs]
I [09:43]: Very good. Um... okay, so that is how the think-aloud process works. And so now, we'll move onto the sample video evaluation. And I realize I've neglected to do something, I'm going to ask you to fill out the same rubric that you normally would, um... why don't you open up your most recent Webassign evaluation just so you can see what it is?
S [10:11]: Okay.
I [10:12]: And while you're watching the video, I'm going to ask you to screencast, that way I can record what's on your screen when you're speaking. Um, you can do that under the... sorry, not the settings menu... the, on the sidebar, the second icon says "screenshare". If you could click on that please.
S [10:31]: Okay, um, let me pull up the Webassign real quick.
I [10:34]: And do bear in mind that once you click that, I will be able to see what's on your screen, so please hide anything you don't, you know... anything with personal information.
S [10:41]: [laughs]
S [10:55]: So basically just do what I usually do with the, um... with all the Webassign guidelines, like organization structure, content models, all that.
I [11:06]: Yep. Yeah, I want you to just look at that rubric and evaluate this video as if you were evaluating any other video you were doing throughout the semester.
S [11:11]: Okay.
I [11:13]: Okay. And it's the video that I attached to the first email.
S [11:15]: Alright, um... let's see...
S [11:28]: Alright, I started the screenshare.
I [11:30]: Okay good, I can see the screenshare, and I can see you have the video up.
S [11:33]: Oops.
I [11:34]: So please watch it and evaluate it as you normally would. Again, I will mostly just be... I will not be prompting you with anything, uh, I might just say "keep talking" if you're silent.
S [11:47]: Okay. Um... do you want me to talk as the video goes, or like, finish the video and then start talking?
I [11:52]: That's—that is rather difficult, I know it's difficult to talk while you're watching videos, I don't expect you to say anything the first time you're going through the video. If something leaps out at you or you get a big idea, please announce it.
S [12:04]: Okay. [plays video]
S [18:06]: Um, okay, I have a question real quick.
I [18:08]: M-hmm.
S [18:10]: So this is supposed to be the same lab that we did, right? The last one?
I [18:13]: Uh, yes, yes, it's the same.
S [18:17]: Okay, I'm going to do the thing now. Um, I like the intro, it explained everything she was going to do in the problem, and...
S [18:24]: Her conclusion was good, too, because it described the what if question. Um... it contains, like, the major—preview of the major sections, and uh, the results also. The only thing I didn't really like was the structure of it, so that's why, like, it's a—I'd give it a very good or an excellent.
I [18:46]: Please forgive me, uh, when you're filling—so, I'd like you to evaluate it now. I realize I didn't send you an interactive rubric. If you could go open up a text document or something and type in your responses there...
S [18:58]: Oh, yeah, sure, hold on. [opens notes.app] Okay, can you see this one.
I [19:05]: Yes.
S [20:07]: Um, second one... let's see. Content models. I would say that was um, let me just do the explanation for this. She identifies [typing] all relevant models... and the main physics principles at the beginning of the video, in the intro.
S [20:38]: Um, doesn't really mention Newton's second law towards the beginning, but she does in the middle, so it's okay I guess. [typing]
S [21:00]: Um, one thing I had a problem with was that the explanation to find k s, spring stiffness, was way too long. Um... [typing]
S [21:15]: I understand that the oscillations mattered, in the final what if question, cuz it asks the difference between the x and y oscillations, and I liked her explanations of that.
S [21:36]: ...but the fact that she did it in the middle of the video just kinda threw off the focus of the whole video, because it was about finding the graph with all the potential and kinetic energies, and not really focused on the oscillations. So I'm going to give that a good.
S [21:51]: Um, for the third one... the data that she used to initialize the model was really identified, and I liked her parameters on how to fit data, it's just that the organization was a bit off because, um...
S [22:10]: Wait. Yeah. [laughing] The organization was just a bit off because of the... of the oscillations instead of the main graph. [typing]
S [22:24]: Because of focus of the video... was... [typing, muttering to self]
S [22:42]: Um... I wish she discussed the code more. Like her reasoning behind the code. [typing]
S [22:59]: And her... reasoning behind the code, especially with the kinetic energy and potential energy formulas.
S [23:14]: Um, I'm going to give that a good, because the explanation was still good, it was just a bit too long for my liking.
S [23:21]: The overall content, um, did not contain any physics errors as far as I could see.
S [23:34]: And explained all her stuff well, except for the kinetic potential and, um, those formulas and the change in those fomulas.
S [23:45]: Uh, umm, I wish she talked more about Newton's second law in relation to this, and when she explained [typing, muttering to self] ...impact...
S [24:07]: And, when she explained about the graphs... [sighs] and the computational model.

S [24:27]: Um, let me give that... a good. The direction of the video was very good. The audio was... good, kind of. It lagged at times but it was good. Um... the code was in parts, and not just like a whole wad of—a whole wall of code, so that's good. And [unintelligible] No distractions on video delivery.
S [24:59]: I'm done.
I [25:02]: Okay I am sending you, uh, the expert evaluation that you would have seen if, uh, this had been shown to you as a practice video. Uh, sending it right now.
S [25:14]: I mean, I always know that his, like... whenever I do the calibrations, I know that the actual grade is much lower than what I give them. But I dunno.
I [25:24]: Could you talk a little more about that?
S [25:26]: Oh, yeah. Whenever I grade—oh, whoops! Whenever I grade the videos, whenever um... like, I look at them. And I always give them a very good, good, or like fair things, and then when I look at the actual, um... the actual grade that the person—the grader gave them, it's always fair or poor.
I [25:54]: Why do you think that is?
S [25:56]: Okay, so this one's pretty close. [looking at expert evaluation]
I [26:04]: While you're comparing those two, if you could please talk aloud.
S [26:09]: Talk about this? The comparison?
I [26:09]: What the impression you're getting about comparing the expert evaluation to your evaluation.
S [26:18]: Um... I mean, it's pretty close. Like the intro was good and I thought it was good too. And the overview is like, it... develops a structure and results very well, um... for number two, for the content models, I did think that she should have spent more time covering the principles. Um... and yeah, she focused way too much on the x and y oscillations and didn't—like, that's why the spring stiffness is there, but she didn't really explain the L oh *[ed: L_0]*, which is also what the actual grader mentions in the third part.
S [26:48]: Um... for the fourth one, okay, [reading aloud] "some sloppiness, mass of the period". I didn't really notice that. It said "mass of the period." "Energy over delta energy confusion, with the..." Oh yeah! She mentioned heat transfer at the end, which I wouldn't really have done. I guess it was really good that she discussed that there was a discrepancy in the data, but I wouldn't have brought it up if I wasn't going to explain it more.
S [27:15]: Um... "missed x y spring pendulum behavior" okay, I also did not notice that it had pendulum behavior. Um, "some audio gets cut off"... oh yeah, that's true. Um... "slides are dense with calculations" [muttering] oh yeah, a lot of ideas could have been expressed... better, I guess. Like, the k calculation could have been discussed less, and the L 0 could have been discussed, of course—the relaxed length could have been discussed.
S [27:45]: Uh, that's mainly it.
I [27:48]: Okay. Well, thank you very much. I'd like to move into some follow-up questions that I have about the experience we've just had.
S [27:58]: Okay.
I [27:59]: So one thing that I noticed you said is that when you were doing these expert evaluations—oh and actually, can you turn off your screencast so I can see you?
S [28:10]: Okay, yeah [does so]
I [28:10]: Thank you. You mentioned very specifically that when you would do these expert evaluations—er, these evaluations, you would see that your scores were higher than the other person's score. Um, now do you have any idea why that might be? Why do you think you were giving higher scores?
S [28:27]: Um... I dunno. I guess I just wouldn't give a poor just because someone mentioned a thing like—failed to mention something. Um, and I'm also not really very harsh grader, also when I was like grading essays in high school and stuff, I never really graded them very harshly. Um... but when some person misses some obvious things, I do give it like a fair or a poor. But overall, I just kind of give it, I just like—I look at their overall delivery, instead of like picking apart little things that they did, I guess.
I [29:00]: Can—and why do you think the experts evaluated this lab—this video, but also the other videos in the way that they did? Why do you think they ended up with lower ratings?
I [29:10]: I mean, I do think they pick up on individual things that we do. So like every little section matters to them. So they grade on every little section, instead of like the whole thing. So I guess they watch the whole video once, and then they like go back through it and watch the individual sections and grade on that.
I [29:27]: Uh, yes, and I noticed that when you were going through the video, you just watched through it once and filled out the evaluation. Is that—
S [29:38]: No, I write down notes sometimes [shows notes]
I [29:38]: Oh, you do?
S [29:38]: Yeah.
I [29:38]: If you could please send me those notes afterward, that would be great.
S [29:42]: I mean, okay, they're terrible, but yeah. It's mostly just like sentences, like... [reading aloud] "Mentions drag force explanation" and "energy principle hashtag good". Stuff like that.
I [29:54]: And do you find yourself that's normally how you're approaching these videos? You normally watch them straight through and take notes? Or do you sometimes go back and check...
S [30:02]: I sometimes go back. When I watch them, I watch them and like take notes in the rubric itself. Because I have that pulled up in another window usually. And then sometimes I have to clarify about like the content models, what they explained in the video, I sometimes go back to just that part and watch it over again.
I [30:18]: Okay. Um, and now as the semester's been going on, and you've been seeing—uh, you've been doing these labs more, do you find yourself—do you think you find yourself giving lower grades?
S [30:29]: Oh yeah, like my grades have like... the first lab, for my lab calibration score, I got a really low score. And I was like, wait, no, I need to go do better on this. So I usually pay more attention to the videos as I'm watching them, and I try to grade them as the grader would grade them, I guess. So yeah [inaudible]
I [30:46]: As the instructor would grade them?
S [30:49]: Yeah.
I [30:49]: And so, with that, how did your experience in the second lab compare to your experience in the first lab?
S [30:59]: I mean, my lab calibration score was higher, so yay. [shrugs]
I [31:02]: Okay, um, and so what sort of expectations did you have for this whole process when you first started doing it?
S [31:11]: Just like the labs, or the online course overall?
I [31:13]: Particularly the labs, but we can talk about the whole class too.
S [31:17]: Um... I didn't think the labs would be that hard.
I [31:20]: Are the labs very hard?
S [31:21]: They're not that hard. But I didn't think they'd be that hard in terms of the grading, like I thought that we would turn it in and instructors would grade them. But, yeah, I didn't expect to have the peer eval grades, like as our main grades, and I didn't expect to have to grade the labs myself. I thought the TAs would—like, in chem, the TAs or professor would grade them. And I guess it was nicer because, I dunno, I got to see what other people's labs were like, and how they like approach the whole problem.
I [31:50]: And what all do you think, um, overall what are your feeling about the fact that you are responsible for grading other people's labs and other people are grading yours?
S [32:00]: I'm okay with it. It was good, I guess. Um, it took up more time than I expected but that's okay. I'm not really doing anything else in the summer.
I [32:13]: Can we talk about that a bit—how long did you think—how long did you expect that it would take?
S [32:15]: Oh, I mean I just—going into the course, I thought that I would just turn in the lab, and that was it, but um... I saw that you have to like... that we got back the lab, and had to grade other people's labs, usually five of them, um... so that took up like, more time than I expected it to take.
I [32:30]: Okay, very good. So is there anything you'd like to talk about in general, your general experience with the whole course?
S [32:36]: Not really, I mean my TA was kind of helpful, more helpful than I expected, because in-class recitations you'd go and like you don't really have to do anything. I guess TAs just go over questions. I mean I didn't really do anything in my other recitation. My only recitation was chem, nevermind. Um... but for this one, it's kind of like, since it's video chat, and I really talked in—my other members of my group, and we all like discussed these problems together, so that helped a lot. More than I expected it to.
I [33:05]: Oh, okay, and so you say it helped more than you expected it to—
S [33:10]: Yeah, I didn't think the recitations would be this involved. Like I thought I'd go in, and like say a few words and then peace out. But um they helped a lot.
I [33:17]: Oh good. So your overall—your expectations had been exceeded—
S [33:22]: Yeah!
I [33:22]: —in terms of the interactions. Very good. And I have one more specific question that I took a note of. You said several times—so going back to when you were just evaluating the video now, when you were describing responses you said several times, you know, "this explanation took way too long. She spent too much time on this." Uh, I'd like to explore that a little bit.
S [33:46]: Okay. So you know how she did the—she did the x and y oscillations quite a bit, and when I was watching the video, I had to like think back on what the main objective of her whole video was, and she kind of just shifted the objective of the video to examining what the graphs at the end, like, the kinetic and potential energy graphs... that was sort of like our main objective, right? To just compare those graphs. And she kind of shifted the objective from comparing those graphs to finding the k_s value. So I... yeah. I guess that was way too long an explanation on the spring stiffness value, and not enough on the main plot that we got in the end.
I [34:25]: Okay, and then... thank you very much. One more question, uh, when you are—okay, so at the very beginning of the course, you said that you found yourself giving higher grades, uh, than the experts. Do the expert evaluations, as you're going through the course, do they seem to make more sense, or do sometimes you'll—there'll be an evaluation, and it'll be a surprise, you wouldn't have expected that part to get that rating.
S [34:53]: Actually, yeah, when I was doing the calibrations last week for lab 3, most of my grades were pretty in-line with what the instructor gave them, but the last one... I didn't think it was that good because, like... I dunno, I didn't think like the explanations were that clear, or the quality wasn't that good, but the instructor actually gave it a higher grade than I did.
I [35:16]: Okay. Because you expected that the instructor grade would be...
S [35:19]: Low. Yeah.
I [35:21]: And what counts as a "low" grade to you?
S [35:23]: I mean, all poors is pretty low. Plus everything that has, like, poor and fair is pretty low, I guess.
I [35:31]: So "fair" is also, you would say, a low grade?
S [35:33]: Yeah.
I [35:35]: Okay. Describe your perception, if you would, of like the rating scale. What does poor mean to you, what does very good mean to you, so...
S [35:44]: I mean, this is just... personally, excellent means just like no fault, perfect all the way through. Very good is just like, it's really good! But it could be better, I guess. Good is [gestures] ehh. Fair is like, I mean, good is like still pretty good, but not like very good. I guess. Um... fair is like, I guess you tried, but it's not all that good. And poor is like, you didn't try at all.
I [36:09]: And, have your perceptions—has your attitude toward that rating, you think, been changing as you've been going through the semester?
S [36:17]: Not really? I guess. Yeah. But my grading of the videos has definitely changed. Like, something that I thought was very good towards the beginning might just be like good now, or fair now.
I [36:30]: So you might think very good, but like, punch in good?
S [36:34]: No. So towards the beginning of the course, when I graded Lab 1 or something, I'd be like "oh, this is really good!" but then like, looking back on it now I'd be, "oh, this wasn't that good", it was just like good.

F.2 Student 2, Spring 2014

Interviewer [03]: Okay, we are recording. Uh, so thank you so much for participating in this interview. Before we begin, I've sent you a consent form that I'd like you to sign, along with a media release form. The consent form tells you the nature of this study—there's no known risk to you for participating in this study—uh, the media release form, if you choose to sign it, gives me permission to share anonymized recordings of this

interview with other researchers for research purposes only. You don't need to sign the media release form to participate in this study, but if you do choose to participate in this stud, you will need to sign the consent form. Uh, do you understand?
Student [00:41]: mm-hmm.
I [00:42]: Okay, excellent. So, if you're reading through the consent form, please take a moment to read it—you can sign it electronically. Are you on a Mac?
S [00:49]: Yes.
I [00:50]: If you open it up in Preview, you should be able to type in your name, or use the signature tool to sign your name to it.
S [00:58]: Okay.
S [01:20]: Okay.
I [01:21]: Okay. And if you could please email that back to me. I need to gather consent before we...
S [01:27]: [nods] Mm-hmm.
I [01:27]: ...conduct the experiment.
I [01:30]: Uh, and while you're doing that, the media release form is slightly different. So, uh... I will be saving the recording of this interview, I'll be using it for my research and for Mike Schatz's—my principal investigator's—research. The media release, form, uh, determines whether we can share any part of these recordings with collaborators or present, for example, clips of videos at research conferences. Um, and so that's completely optional. You can choose the level of anonymity with which you are comfortable, um... and, uh, please... you don't actually—you can in fact wait to send that to me until we've finished the interview.
S [02:08]: Uh-huh.
I [02:09]: Okay, so, let's go over what we're doing today. I'd like to begin by walking you through a think-aloud problem for practice, then I'll ask you to evaluate a sample lab video—the video I've attached in that email—while you think aloud, just as if you were evaluating it for your lab practice and calibration exercises. My goal is to understand what you think about the video, and why you're giving the ratings and the comments that you do. Next, we'll discuss that experiment, and your experience in the course so far, and then we'll end with some follow-up question. I—we expect the whole interview to last about an hour. I'll be recording this interview over the webcam—I'm currently recording. And when you're doing your evaluation, uh, I will ask to record your computer screen so that I can see what parts of the video you are watching. Uh, so, with that understood, would you like to begin?
S [03:04]: Sure.
I [03:05]: Excellent. Okay, so this—part of this study will involve a think-aloud section, and the way a think-aloud section works is that... uh, when you're solving a problem or evaluating a video, I want you to narrate to me the thoughts that you're having as you solve them. The thoughts don't have to be coherent, they don't have to be in complete sentences. I will not as you any further questions, I will mostly remain silent during that part of the interview, but it—but if you're quiet for a little while, I may prompt you from time to time with "keep talking."
S [03:44]: [nods] Mm-hmm.
I [03:45]: So does that make sense and would you like to begin?
S [03:47]: [interrupting] Mm-hmm.
I [03:48]: Okay, so, the question that we'll use to warm up is a two-part estimation question. Okay. What is the volume of the room you're in right now, and if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for the room to fill with water?
S [04:07]: [looking around] Mmmmm...
I [04:13]: Keep talking, please.
S [04:15]: Okay, I'm just trying to like estimate probably like... [gestures] the length of the walls and then the height of the room. Um... so I would say probably...
S [04:27]: The length of one wall is probably like... eight meters. And the other wall's probably like... ten meters. And then the height I would say like... seven meters. So.. that'd be... eighty times seven. So that would be... my brain's going blank. [laughs]
S [04:59]: Five hundred and sixty. Meters. ...cubed. [laughs]
I [05:05]: And the second part of the question is, if you were to turn on an ordinary kitchen faucet in that room, about how long would it take to fill with water—and you may use a calculator or a piece of paper if you like.
S [05:18]: Okay. [reaches for something]
S [05:23]: [looks around] Um, I'm trying to think I guess of like how fast... the water would be running. To figure that out, um... [looks down, types on calculator?]
I [05:43]: Keep talking please.
S [05:45]: Okay, um... I just used gravity as, like, how fast the acceleration of the water was, so when... that would be... five thousand—five thousand five hundred seconds. About.
I [06:05]: Okay. Thank you very much. That's generally how a think aloud works. You did very well, I will just occasionally prompt you to keep talking. Um, and so now we will move onto the sample video evaluation. So, uh, in that email I sent you, I sent you a video to watch and just for reference, a copy of the rubric that you've been using to evaluate videos so far. Uh, so I'll show you the expert responses to that video after you finish evaluating it, much like you do in your practice and calibration assignments, and of course this is for research only. Your evaluation will not affect any of the grades in your courses.
S [06:31]: [interrupting twice] mm-hmm
I [06:42]: Uh, okay, so just like last time I'd like you to narrate to me the thoughts going through your head. And while you're viewing the video, if you could please share your screen with me. So that I can see—and, uh, bear in mind as soon as you click the share screen button I can see everything on your screen, so if you've—personal information, you should close it.
I [07:20]: Okay.
I [07:23]: Alright. So then, just as I'm watching, like, say how I would like grade on the rubric?
I [07:30]: Yeah, so, when you're watching the video it's going to be difficult to narrate your thoughts because you're paying attention to the video.
S [07:36]: [interrupting] mm-hmm
I [07:36]: Uh, but when you are doing your evaluations when... I would like you to like open a text document a rating an eval—and a comment. When you're doing that, please narrate to me what you're thinking.
S [07:51]: Okay.
I [07:52]: While you're watching the video, if anything—if you have a strong thought, if something leaps out to you, please mention that.
S [07:58]: Okay.
I [08:00]: Okay, ready to begin whenever you are.
S [08:02]: Okay.
I [14:29]: Keep talking, please.
S [14:30]: Okay. Um... let me... [opens rubric] It's kind of hard to see the rubric, it's kind of fuzzy, but I'm going to like try to do from what I remember it being...
I [14:45]: Huh, actually, no, I will send you a better version.
S [14:48]: [laughs] Okay.
I [14:50]: It is important to be able read it. Huh. I am very sorry.
S [14:53]: Oh no, it's fine.
I [14:55]: [muttering] ...low resolution... ...what's the best way to send this to you...
S [15:11]: Mmmm...
I [15:14]: I'll, uh, I'll send it in an email by itself.
S [15:18]: Okay.
I [15:19]: I think it scaled itself down.
S [15:20]: Yeah it just is a really small image.
I [15:23]: Yeah.
I [15:43]: Okay I've just sent it a second time. Um... when it comes in, please just... take it and then we can continue the think-aloud study.
S [15:59]: Okay. So. Um, so, for the first section of the rubric, um, the organization structure, I thought that it was... either between very good and excellent, um. There was a really clear organization with the numbers of all the different, uh, parts of the lab that they needed to consider. Um, and she introduced the problem and like the fundamental principle she was gonna use. And then she also briefly stated the result, and it was all clearly laid out on the slides, so even if you missed a little of it, because it was a little fast-paced, you could still see what she was talking about.
S [16:48]: So I would probably give that an excellent. On the rubric.
S [16:54]: And then, for the content models, um... she talked about both newton's second law, um, and then later on the energy principle that she used in the lab. Um... the only thing that I thought was like a little hard to follow was because the parts of the code were so split up between what she was talking about. It was kind of hard to see like the sequence of the code, and how it was actually um, written out, because [switches to video window] she had like the code here, and then she went to here where she just started talking about the data again, so I think maybe like... putting this before that would have been helpful, um, to follow along.
S [17:41]: So I would say that was probably good.
S [17:46]: Um, and then for the prediction discussion, um, she clearly talked about how she got her data, especially here, um, she tried both um periods to see which one was better in her model. Um, and then she also talked about how she used the data in the csv file here, um, to put into her code. Um, she also discussed how her data um, differed from the model. Um, between the two different situations and why that might have been, and she talked about that again later on, near the end of the video.
S [18:30]: So I would say that was very good.
S [18:32]: Um, and then, I didn't see any major physics errors. Um, and I think she answered both of the... um, what if question, and the what does it mean question when she talked about, um, does the energy principle hold for this. Um... experiment. And then, and here she talked about how it did because, um, that the total en—total change in energy was zero, which was what it was asking us to look at.
S [19:04]: And then she also compared the x and the y oscillations. And talked about which better fit her data... um, with her model.
S [19:16]: And then I thought it was a very well put-together video, um, it was either—oh, so, for the content, I would give it a... very... er... I'd give an excellent.
S [19:27]: And then for the production delivery, I thought it was really well put-together. Things were labeled, like, clearly, um... she talked a little fast, but I think there was just a lot of stuff to fit in. So maybe summarize a few things a little bit. Um, and maybe adjust the order things are presented, because sometimes I thought that it was like backtracking to something else instead of like a smooth flow of like things were set up in the code.
S [19:57]: So I would give that a very good.
I [20:03]: Okay, I am sending you now—so this is as if you were—is this the evaluation you would submit for your calibration exercise?
S [20:12]: Yes.
I [20:12]: Excellent, okay, I am now sending you the um... instructor ratings for that video.
S [20:28]: Okay.
I [20:28]: Okay. And, now please, we'll continue the think-aloud study if, uh, if normally you would compare your evaluation to the expert evaluations, please do that.
S [20:39]: Okay.
S [20:41]: Um... so I said like for the first part, um, I gave an excellent and they gave a very good. I think that this is one thing I like struggle with a little bit, in the actual um... like, in the actual labs that we've done. I think I find it hard to like... it—I think it's easy to differentiate like poor—like if something's poor I can like easily say, like "Oh that was poor, maybe fair." But I think it's hard to differentiate between very good and excellent and good, because I tend to notice that like... in the ones we've done, a lot of the calibration labs never have excellent.
S [21:22]: So I don't know, like, what constitutes an excellent from a very good. I guess is what—it's hard for me personally. Um...
S [21:32]: And then [reads screen]... I do agree with the second part, that she could have talked about the principles more, because, um... she kind of just mentioned them, but if you were somebody who wasn't really familiar with the topic, it might have been a little confusing.
S [21:56]: Uh...
I [22:13]: Keep talking please.
S [22:15]: Um, I pretty much agree with like all the things that this... expert rubric is saying. Um, I think there's some things that I like missed. Um... when I was grading mine.
S [22:30]: But I think they're all things that, like, I agree with, when I like look back on it.
S [22:39]: Um... and I think a lot of like, the things that I like, I guess maybe the discrepancies between this two, are that um... I think there was like some things written out, like she wrote out how she like, the L nought equation at the bottom of one of the

slides, um, but she never mentioned it, so I think like maybe I saw that and I was like "Oh! Well it said where the L nought came from." but really, the video didn't discuss that, so I like... I think that it's true that she should have talked about that.
S [23:16]: But other than that, I think like... it was like fairly similar to what I was getting at.
I [23:25]: Okay, thank you very much, that was a very productive exercise.
I [23:30]: And so now we will go into our, uh, interview and follow up session.
I [23:38]: Okay, so first I'd like to talk about some of the—some things I noticed when you were describing how you evaluated these videos and what you thought about the video. So, you—I noticed that when you were evaluating it, you just sort of played the video straight through. Is that normally how you would do it?
S [23:56]: Yeah, I ususally—I'll play it straight through and either like take some notes like I was doing on the side, or like, sometimes I just like remember basic like main points. Um, which I think I've gotten better—or like, easier as I've done more of them because I started to like, know what the certain parts of the rubric are. And then usually what I do is after I've watched it, I kind of know like certain sections of the video and where they are. So then I go through the rubric and I kind of go back to that section and rewatch it, or either just click through it paused. To like refresh my memory exactly on what like they were doing in that section.
I [24:39]: Okay. And, uh, has that... um—so we're now pretty well, we're toward the end of the semester, have you developed those skills? Has that become easier, or were you not sure how to do it at the beginning?
S [24:53]: I think, um, I kind of tried both ways of like, I tried once doing it while I was watching it. Like, completely grading it while I was watching it. I think I tried that the first time. And I didn't think it was, like, very helpful, because I was getting overwhelmed by trying to type comments and then like watch and type comments, and then I'd miss things. And I think like even just like missing one sentence of the video could like cost that person's grade. [laughs]
S [25:23]: So I think that how how I ended up doing it later is like, better for me.
I [25:29]: So, that—that seemed to me like a powerful phrase: "missing one sentence could cost someone else their grade". What's your—what do you think about other people's grades?
S [25:40]: Um, well I guess like I also think about somebody else is doing that for me, so like, I would hope that they were paying attention to my video, and not just like giving me random grades, cuz they missed stuff. And I noticed like in the ones I did do at the same time while grading, I would like, think like, "Oh, I didn't hear them ever mention that, like, principle, or that part of the lab." And then I'd go back and realize that I was just typing something while they said that.
I [26:11]: Um... do you think other people in your class have that same attitude toward you? So, you hope they do, but do you think they do?
S [26:19]: I feel like they do. I mean like, and also like I guess based on having my grades back, I feel like they're grades that I'm okay with. I feel like they accurately represent, like, the work I've done.
I [26:34]: Did you expect that that would be the case?
S [26:37]: Um, I think so, yeah. I think, like, I dunno, maybe this is just like my optimism, but I guess like I think that... it'd be different—like, I guess I think that if people know that maybe if someone else is grading them, too, that like... they have the same risks involved, so like... they know someone's grading theirs fairly, hopefully, so they should grade someone else's fairly.
I [27:09]: Very good. Um, okay, so, you also mentioned a while ago that it's uh... you don't think it's very easy to distinguish excellent from very good. Could we talk about that for a bit?
S [27:24]: Well, I just noticed, like... I think it's interesting I guess when I do the calibrations, if there's a video that's like poor all the way through, or poor and maybe a little fair, I usually get that like completely accurate to what the like expert, um, rubric is. When I look at the answers. But then when there's one that's like, either has like between excellent or like very good answers, that's where I tend to like get off from that one. So, I don't know why it's harder for me, I guess, to differentiate between the higher values than the lower values. I guess it's kind of like—I guess when I think about it, like, to me it's easy to say, "Oh, well that doesn't have that, oh like they didn't mention that at all"... more... it's easier to do that than to like say "Oh, did they mention that enough?" or like "Was there enough emphasis on this?" kind of thing.
I [28:22]: And have you... so do you think... do you think your attitude toward that has changed throughout the semester?
S [28:30]: I think I've gotten better at it. Like, as I think I've kind of—as the calibrations have gone one, I've kind of like, gotten used to what I think the expert, like, considers excellent and very good. But sometimes, like, I guess sometimes what I think I do when I do the calibration is like, I might think it's excellent, but I'm like, "Oh, they probably thought it was very good." So then I'm like, "Oh, it's probably very good". [laughs]
I [28:55]: "They" being the experts who fill out those evaluations?
S [29:01]: [nods emphatically]
I [29:01]: Oh, so I would like to talk about that topic a bit. You said very boldly earlier that they almost never have excellents. The expert scores almost never have excellents. What do you think about that?
S [29:14]: Um... I dunno. I guess maybe it has to do with like, I... don't think anyone's video is like perfect. Um... so probably it's very hard to get an excellent. Um. And I also—I guess, too, like... for me, like, I'm not always going to be right on what I think. So like, I—there might have been a physics error in somebody's video that, like, I thought it was right too. So, I mean, like, I'm still in the same level of student as they are. So I think that... I might be missing stuff, or like, think stuff's really good when there's actually... a slight error or something.
I [30:01]: Okay. Um... so... on that note, you're—you're describing the experts having a, different set of knowledge about physics. A larger—more knowledge about physics, a better ability to attend to physics details. But overall, if you had to describe it generally, what do you think is the difference between how the experts and how you evaluate these videos?
S [30:36]: Um, I think... like, I think now, I would say I'm a lot... like... more accurate, I guess, in terms of like comparing mine towards the experts. Um, I think the first time I did it, mine were, like—I think I was like being very nice, and I gave a lot of excellents. And then I realized, like, "Oh, that was like not—not what it was at all" [laughs] So I think over time I've realized to be—to be more critical toward the video, I guess. Cuz at first I was like "Oh, yeah, that was good, that was good!" And then like... as I've gone on, I've like gotten more used to like... being very critical towards like certain things that people say or like... I guess, like, if people say something wrong, it's not—I think in the beginning, I was like, "Oh, well, maybe they just like said the wrong thing." But like... they said the wrong thing, so I should mark them down.
I [31:36]: So at the beginning you said you thought—at the beginning of the semester you, uh, felt like you were being nice to your fellow students. Why did you feel like being nice to your fellow students?
S [31:47]: Well, I also—I think I was like... not sure how harsh to be. [laughs] I guess! Especially before I did—like the first ever calibration video, I didn't know like... yeah, we had the rubric, but I didn't know like... how direct to follow the rubric. Um, especially because there's only like two ends of the spectrum listed. So like, I think a lot of times I'd be like, "Oh, well, they had this and this, so maybe I'll give them a good or a very good." But really, they should have had like more points than like what they had to get a higher-up grade. Um, so I don't know if it was more me being nice or me like, not knowing... what was expected, I guess.
I [32:35]: Uh, and... uh, as a corollary to that, you said you were being nice earlier, now you're being more critical. Does that imply to you that being critical is mean?
S [32:50]: I guess not. I think, it's just like... I mean cuz I'm not... I guess when I'm being critical, I'm not like purposefully trying to like beat that person down, or like give them a bad grade. Like, I don't want to give anyone a bad grade! But, like, I think it's important to give them the grade that the video deserves, which takes you being critical.
I [33:18]: I like that. And so just a moment ago, you talked about how the rubric didn't have all that much information, or didn't have enough information, uh, at the beginning of the semester. Could you talk about that a little bit more?
S [33:31]: Mm-hmmm. Well, I'm a person who's like very bad about making decisions, and like "Oh! Maybe it's this, maybe it's that." And so I think if there's not like concrete, "it must have this", I'm like, "Oh, well I don't know if it's"—like, should it have half of them, should it have three-quarters of them to be very good? So I think like—and that might be also why it was like easier to like figure out the poor videos, cuz I had a list of like what made that video poor, like "Oh, it didn't have this, this, this, and this."
S [34:06]: And so I knew like, "Oh! Well, it fit the exact description of poor for that section." But then when you got to the middle three, it was like, I kind of had to interpret what I thought, like, how many bullet points made it very good, or how many bullet points made it good.
I [34:26]: Okay. Um... let's see... and can you just generally describe to me the first... your attitude and your general experience when you were first asked to evaluate your peers' videos.
S [34:50]: Um... at first, like... I like, I guess to me, I thought like... "Oh, they're probably all going to be the same." Like it might be hard to like grade them. But after doing it, like, they're all like very different, which is kind of interesting to me, because, like, we're all taking the same class. Yes, we have like different TAs that we meet with and stuff, but we follow the same exact rubric, like, we have the same exact lab. But it's interesting like... what people focus on in their videos, and like, how do people set up their videos. And I think that helped me like make my videos, too, because like, the first time I didn't use slides at all in my video. I just like talked about the video—about the code, and about my lab, and then after watching the first set of them, I like realized how helpful it was to like... use the powerpoint and speak over that, and so I think like... grading people's also helped me improve my own video.
I [35:57]: And what are some of the "interesting" things that other people focused on in their videos?
S [36:01]: Um, I thin, cuz like I guess for me, I tend to focus a lot on the code, and like... what goes into the code, and how I set it up. And some people kind of just skim over that, and talk about the data instead of the code. And so I like think it's interesting interpret the directions differently, I guess. Like... there's kind of two parts of the lab, the data and the code, and some people tend to focus a lot on the data and some people tend to focus solely on the code. And so I think it's interesting that like, I don't know if it's different people's like... knowledge about the different ones, or like... what they want to talk about, but it's interesting how like watching the five videos for the like, um, to grade the labs, they're all like very different to me.
I [36:50]: And why do you think you pay so much attention to the code in your own videos?
S [36:55]: Um... I guess for me, I like... I guess it's more of like, how much time I devote to that when doing the lab. So for me, I guess, like... yeah, I take the video and then I put it in Tracker, but like to me, that was like a minor part of the lab, and I spent more time working on like, setting up the code, and running—and like making sure the code was correct. So I think that I guess in my mind, it's like, that to me was the focus of the lab. And not like, the actual like, getting of the data. And like tracking the data.
I [37:33]: Uh, well, thank you very much for your time. I think that was a very successful interview. I need a couple more things from you. I need that media release form, answered as you like, and the other thing I need, you'll find in the email a payment form, so we'll be giving you twenty dollars for your time here. Uh, if you're—I can't hand you, you know, money through here [gestures at computer screen], so if you're a GT employee, like you're on work study or something, you can give me your gtID and I'll give you twenty bucks on your next paycheck. Otherwise we can mail you a paper check.
S [38:07]: Okay.
I [38:08]: Send both of those to me as soon as possible.
S [38:11]: Mm-hmm.
I [38:12]: And thank you again for your time, I hope you have a great day!
S [38:14]: Thank you.
I [38:16]: Okay.

F.3 Student 3, Summer 2016

Interviewer [04.19]: Okay, excellent, we're live. Alright. Thank you very much for agreeing to participate in this study today. I've sent you a consent form that I'd like you to sign, along with a media release form in that email you received earlier today. The consent form describes the nature of this study, and that there's no known risk to you for participating. The media release form is separate, so if you choose to sign the

media release form, you give me permission to share anonymous recordings of this interview with other researchers besides my principal investigator Mike Schatz for research purposes only. Uh, so you don't need to sign the media release form to participate in this study, but if you do choose to participate in this study then you do need to sign the consent form. Uh, do you understand?
Student 47.01]: Yep! I will... uh, can I sign those and send them to you afterwards?
I [51]: You can sign the media release form after we're done recording, uh, but I actually do need to gather consent before we begin the study. So if you could please fill out the consent form now.
S [01:02]: Oh, okay.
I [01:03]: Uh, are you on a Mac?
S [01:05]: I am.
I [01:06]: So, you should be able to either use the signature... like, open up the pdf in Preview, and you can either use the signature tool or type in your name and date. And that will count as your signature.
S [01:17]: Okay.
S [01:53]: [inaudible]
S [03:08]: Okay. I'm saving it now and sending it.
I [03:11]: Excellent, thank you.
I [03:19]: Yeah. So the way this works is the consent form allows me to conduct research. It allows us to conduct this study. And the recording of this study, uh, will be viewed of course by me and by my principle investigator Mike Schatz. For research purposes only, you know, there will be no influence on your course grades. The media release form is separate and in fact you can think about what we've recorded today and then decide whether you would like that footage to be released or not. And so, I will not be releasing it for... our intent is not to use it commercially, rather you're signing a release for research use of these videos. Research use would be, uh, if we wanted to present a short clip of the interview at a research symposium, or something like that.
S [03:59]: Okay.
I [04:00]: And that's entirely optional. If you don't sign the media release form, we will never release any segment of this video or audio. If you do sign it, you can choose the level of anonymity that you want. So for example, we could... you will be anonymous, your identity will be protected, but you can either choose to... if you're fine with your face and voice being shown, you can choose that. If you would rather have your face obscured or your voice garbled, you can choose that as well.
S [04:27]: Okay.
I [04:28]: Okay. Excellent. Uh, so, please send that to me as soon as we are done with the interview, but for now let's go ahead and get started.
I [04:36]: Okay, so part of this experiment... part of this study is going to involve a "think-aloud" study. So I'll begin by walking you through a think-aloud problem for practice. Uh, then I'll ask you to evaluate a sample lab video for me while you think aloud, and I want you to watch and evaluate this lab video just as if you were doing your practice or calibration assignments like you've been doing this semester. Once you're done typing up an evaluation, you will then get to see the expert evaluation, and again, I'd like you to think aloud while you're comparing your evaluation to the experts'. Uh, finally, we'll have a brief interview period and follow-up questions. And, uh, I said that this would last about an hour—so far they've really been lasting about forty-five minutes.
S [05:24]: Okay.
I [05:25]: And with that, would you like to begin?
S [05:27]: Yeah! No problem.
I [05:29]: Okay, excellent, so first we're going to warm up with a practice physics question so we can get used to how a think-aloud sproblem works. So the way a think aloud study works is that I ask you to perform a task or think about something, and you narrate to me the thoughts that you have in your head as you're performing the task. Um, you... the thoughts don't have to be complete, they do not have to be in complete sentences, they don't have to be particularly coherent. I'm not going to ask you questions, and I'm not going to say very much, except if you're not talking for a little while, I might gently prompt you to keep talking please. Uh, so... to get practice with the way that works, we're going to warm up with a two-part estimation question.
I [06:16]: What is the volume of the room you are in right now, and if you were to take an ordinary kitchen faucet and turn it on, about how long would it take for that room to fill with water?
S [06:29]: Okay... so first of all I guess I'm thinking of the dimensions of the room? And trying to see, like... I would say... maybe about—well, it's a square room, and it might be about a meter... two meters... like... squared, like, two times two, four meters squared. No, a little bit more. I'd say like nine square meters, maybe, of area. And like, base, and then height, I guess that would be maybe... also three meters. So 27 cubed meters. And then if I turn on the faucet, I guess...
S [07:14]: I'm wondering how fast, or how much water comes out of the faucet. I guess? Um... But, I mean, taking into account a normal faucet, I'd say... it would take... maybe four to five hours? To fill. Yeah. [laughs]
I [07:35]: Okay, thank you very much, that was very successful. So that's how the think-aloud study works. And I'll ask you to do the same thing while you're evaluating the lab video that I just emailed you earlier today.
S [07:47]: Mm-hmmm.
I [07:49]: So please go ahead and open that, and while you're watching and evaluating the video, I would like you to screencast your screen so I can follow along with what you're looking at. So please go ahead and switch to screencast, but before you do, bear in mind that I will... as soon as you click the button, I can see everything on your screen, so please hide any personal information.
S [08:09]: Okay. I think I can actually choose to just show the video screen.
I [08:15]: Uh, yes you can, but actually, while you're watching the video, I would like you to, um, I'm going to ask you to type up your evaluation...
S [08:26]: Oh, okay.
I [08:26]: So I would like to see all of those windows.
S [08:28]: Okay...
I [08:35]: And in your email you will also find, just for reference, a copy of the rubic that you've been using so far. Actually, we had a problem with that last time, I need to email you a higher-resolution version.
I [08:49]: Okay.
I [09:01]: Okay, I've sent it.
S [09:19]: Okay. Why is it... oh, okay.
S [10:03]: Okay, can you see that?
I [10:04]: Yes, yes I can see that just fine.
S [10:07]: Okay.
I [10:08]: Okay, if you don't mind, would you mind hiding your message windows? Just in case something comes in.
I [10:16]: And finally, when you're ready to actually fill out your evaluation, you'll need to do that in a blank text document.
S [10:24]: Uhh... okay. So you don't want me to fill it in the table..
I [10:31]: Well you can if you want but it's just an image, you would have to like, use text boxes and things.
S [10:36]: Okay. That seems easier.
S [11:07]: Okay. So... um... they, the video's presented at first, it seems pretty organized, because it has the video, it has like an order of... like, how the person thought of the process. So I like the type of organization. Let me just...
S [12:36]: Okay, so, the... all of the parts of the introduction are included. Um... let's see, what else... "Excellent use..." Okay, so, for now, I would say that is a really good structure, let's see how she goes on.
S [13:46]: Just a general thought... she's talking a little bit too fast. Um... let's see, that would go under production.
S [14:16]: I'll add that later.
S [14:58]: Okay, so she clearly has a really good idea of how... of the physics behind this experiment, so...
S [16:20]: And also, she... she clearly, um, identifies what is being... what the experiment is.
S [17:01]: Okay. Let's just add that.
S [17:39]: Okay, so she actually—differently from the way I thought of it, um, she included the what if, or, what does it mean questions, she included those in-video. I included them only at the end of the video.
S [18:41]: Okay, so another difference is, um, she actually graphed both the x position versus time, and the y position versus time. Which makes the things clearer, I like that.
S [19:40]: Okay, so...
I [19:45]: So before you keep going, I've noticed that the text box method seems to be rather inconvenient. If you would rather just type out a list in, you know, TextEdit or something like that, that would be perfectly fine.
S [19:56]: Okay. I'll just do it in Notes, I guess.
S [20:14]: Okay, so additionally... she covers the errors... between um, the experimental and actual—and theoretical data. Um, and the reasons why those errors exist, which are pretty similar to what I had. Um... she continues to have really good organization.
S [23:12]: Okay, so in general, I guess... I'm not really... I'm not really sure of how much I like the end, like the way she ends the video, I guess, because it's a little bit too fast and I feel like she could have focused a little bit more on explaining the energy principle a little bit better. Although she did go through a good overview of it, but because it's the main part of the experiment, she could have slowed down a little.
S [24:36]: Okay. Um, let's see. So I've already gone through organization, intro, conclusion... states problem, okay. Okay, initialization data, um... [muttering] how data found in Tracker is used in code.
S [25:42]: The follow-up questions... Um, and the video has good resolution. I would just advise to move the mouse a little bit less so it's not distracting.
S [26:17]: Okay, so, um, do you want me to actually grade this right now? [Interviewer]?
I [26:24]: Yes, please.
S [26:30]: Um, okay, I would say organization structure, I would say... she has a logical structure, but her conclusion isn't that... it's good but it's not as good, and it's a little too fast. Uh, it restates the problem. Interaction... transitions... okay, so I would actually, uh mark this as very good.
S [27:05]: Um, content models. Models relevant to the physical system... discussion of main physics ideas, they're good, they're pretty good but they're not fully like 100%. But they're pretty good. Application of ideas to problem... yeah, and excellent connection between fundamental physics principles. Yeah, she clearly mentions the energy principle and how everything connects to what we're learning in class. Um... the—she does go over how to initialize the model with the data. Um, she goes over the errors.
S [27:54]: I would say this is very good—well, excellent, because she does everything well.
S [28:05]: She answers these questions, but they're a little bit too lost in between... yeah, they're not as deep as they should be. So, I would go with good. And...
S [28:23]: She has a mouse with the circle around it, which makes it distracting whenever she moves it. Um, and talks a little too fast. I would say that's good. Um, yeah.
S [28:41]: That's it. [laughs]
I [28:44]: Okay, thank you very much. Um, as soon as we're done, if you could please send me that rubric and the notes you've typed up.
S [28:51]: Okay.
I [28:52]: In the meantime, this... I am sending you right now the expert evaluation for that same video.
S [29:06]: Okay! So they're not... I mean, I guess the general idea that the video's good seems to be... good, I guess. I was a little bit less strict [laughs] than the professional one. um, let's see... very good results in intro with very good review—overview, yeah, that's the same thing I said, basically.
S [29:36]: Um... should have spent more time covering principles' connection between principles and motion... I mean, yeah. I guess I can see that. Um... where did the L sub zero data come from? Good k calculation discussion. Okay! Yeah, I guess in that sense, I did give her a pretty high grade. Um...
S [30:06]: I guess I just... it—I saw the k calculation, but I guess that makes sense, where did the L nought data come from? That... it wasn't actually there. Like, it wasn't well-explained. Um...
S [30:22]: Some sloppiness in language. "Mass of the period". E over delta E confusion without a very clear connection to motion. Unclear discussion of heat transfer.
S [30:36]: Okay... um... they said good, I said good. Um, but yeah I agree, it was a little confusing and the questions weren't really stated out, so... it made it a little bit, not clear. Um, and finally, some audio gets cut off, slides a bit dense with calculation. Headlong pace of narration, many ideas could have been expressed more succintly—yeah. Which is the same thing I thought. She talked a little too fast, and went over some ideas way too fast which were, like, the most important in the video. Um, but overall it was good. So, yeah I guess I gave her a better grade than the... special person did. Maybe because

I went through the, like, I went through the process of making the videos, I understand that, like, making—making that video five minutes isn't as easy it seemed at first. Five minutes seems a lot, and then it comes out to be not as much time you need to explain the data, I guess. Or maybe—yeah.
S [31:48]: I would just say that. Okay.
I [31:52]: Excellent, thank you very much. Um, so now I have, uh, some follow-up questions that I'd like to spend some time discussing, from what I have seen today. Um, so firstly, would you say—so when you were watching the video, you would watch a little bit, stop it, fill something out on the rubric, and then go back to the video and continue. Would you say that that's normally how you view these videos?
S [32:17]: I do. Because I feel like, as—like, if I watch the whole video through, without stopping, I start to forget little details. Um, and so I like to stop and like to start writing down. Plus, sometimes it can be a little bit distracting, like... because we have to evaluate five videos of the same topic, um, it's really easy to like miss some data. Um, and so that's why I like to stop and like, try to like fill in what I can as it goes by.
I [32:51]: And why is it so important not to miss anything?
S [32:56]: Well, because we are grading someone else's work, and they took time just like I took time to do—to make my video. Well, they took time to grade—like to make their own video, and they're also grading my video, so I would want them spending some time, like, taking into account the details I put into it. So. Yeah.
I [33:17]: Well, and on that note, you've said that you feel like you were, uh... you said that you understand it is how hard it is to make a good video like this, and that maybe that's why you—you give the grades that you do. Could you talk a little bit about that?
S [33:37]: Yeah! Um... I'm gonna—can I stop...
I [33:47]: Oh yes, yes, please, I would like to see your reaction.
S [33:51]: Okay, so, yeah I guess... like, um... at first when I saw that we had to make five minute videos, I was like that's way too much time. But then when we actually realized everything you had to include in those five minutes, it's not that much time, and you really have to like stop and organize yourself. Um...
S [34:11]: And I feel like, maybe because—first of all, I'm not a physics professor, so although I get the physics ideas and I'm learning the material, I don't get it as deeply as a physics professor would.
I [34:27]: M-hmm.
S [34:28]: And maybe that's also why like, the way—if someone explains it in a way that's simpler than the way a physics professor would explain it, I understand that explanation. And maybe that's why like, I don't grade as harshly, maybe? Um... not saying that physics professors grade that harshly. But, um...
I [34:50]: So I would like to, um... you did bring up something interesting right there. Do you think—do you find it easier to understand the explanations from your peers than you do from your professors?
S [35:01]: Sometimes. I mean... if I'm learning like, some material for the first time, um... I find it easier to understand from a professor. But then when, like, I guess what I meant was that if I hear someone else, like, my age explaining something in a way that's not as deep as a physics professor would do it, I understand it as well. So, like, I understand even if they're not saying everything, I understand what they're trying to say. Which I guess like if you're grading a video, you're technically supposed to, like, try to like pretend that you don't know as much. And you're trying to learn from zero, from scratch. Because that's like what you're instructed to do with these lab—like we're supposed to pretend that like our viewer doesn't know anything. That we're explaining from scratch.
S [35:58]: But I guess that's just a human error. Like, I just, like, go from what I know and try to understand how the other person is trying to explain it, so... yeah.
I [36:09]: And before we talked about this, you were in the middle of expressing a thought about your strictness in grading, relative to the professors.
S [36:18]: Okay, so yeah I think that's also connected to what I just said, in the sense that like by knowing that the other person isn't a professor or like knowing that they're in the same position I was, where you don't have much time to explain as deeply as you would want to explain it. Um... they're still doing a pretty good job at doing what they are. Although, like I have had to grade, um, some videos where like this person really didn't take the time to at least read what they're going to be graded on. Because it's way too... um, it's—[struggling]—it's an explanation that's like too wide. It wasn't as in-detail as it should be. Um... so I mean, I like to see like, when I see a video, I'm like okay, let's start from how that person's started—at least started off like the video I saw that—it looked organized, it looks like the person took the time, and I'm like okay, this person is actually taking the time to do it.
S [37:28]: But there's some videos where, like, it's pretty obvious that they just did it because they had to, and because they're trying to get—some sort of grade. But like, it's not... it's not that good [laughs]. So...
I [37:42]: So would you say that—what do you think about the... the importance of the overall impression that you get from a video. Because at the very beginning, when you first saw the expert grades, you said, "Oh, the general idea was good. It looks like it got a good overall".
S [37:58]: Yeah. I'll say, an overall, like... I believe in like first ex—like, first impressions. So like, I'm not—like, I don't know the people who I'm grading, all I have to like, I can only base myself on the video that's being shown. So, like—if a person like, put the work into it, like you can see it from the start. And that means like that they actually took the time to like think about it and like work through it, and so like I'm going to appreciate the time that they took to do it. Like, I try to take that time with my videos as well. Um... but if, like, if from the start there's like no introduction whatsoever, there's absolutely like nothing—they just like rush into the video, of Tracker, whatever... like, um I would say first impression is pretty important.
S [38:55]: Because you can tell from the start whether or not the person actually took the time.
I [39:00]: Okay. Um, and do you think experts have that same attitude toward the videos.
S [39:07]: I feel like, in a way, they do. Because I mean, part of the rubric that professors, like they made, like... it's actually, like, is it, does it have like a proper introduction? Does it talk about how—like, the parts of the videos, that it's gonna have, um... or is it just like... just like all over the place. So I feel like in a way, they do. Um, but also they have to take into—like, they obviously take into consideration the physics part of it. A lot more than, like, a student would, because they actually know more about it, they actually dedicate their lives to this, so [laughs].
I [39:49]: So it might have some sort of more, personal importance to an expert.
S [39:54]: Yeah.
I [39:56]: Um. Okay. So you say—you're talking about how you can sympathize with your peers, and especially you can understand how difficult it is to make a video five minutes, how much time they have to put into it. Um, and that therefore you are less strict in your grading than the instructors. Do you think the instructors don't—uh, what sort of attidue do you think the instructors have toward the difficulty and time students put into these videos?
S [40:25]: What do you mean?
I [40:26]: So if you say that you are, um, that you appreciate the amount of time that students—that your peers put into your videos, and therefore you're less strict than the instructors—are you suggesting that the instructors don't appreciate the amount of time that students put un, and are therefore more strict.
S [40:47]: No, no, that—I guess what I mean to say is like, I do it in an unconscious way. Like, I don't actually sit down and be like, okay, that person obviously didn't put any work into it. Like, I'm going to give them a bad grade because of it. No, I mean... I feel like a professor also looks at like the student's work, and I say this out of experience like of whenever I've had like trouble with a professor, or trouble with the subject, like I come up to them with some work done, like, I can see the difference between me coming up with some work done and someone else who literally comes in and be like, I didn't understand absolutely everything—anything, like explain everything all over again. Like, it's... you can tell when a professor, like, that a professor likes obviously appreciates that.
S [41:39]: I guess... the difference between grading comes down to like, um... I'm a student, I don't know as much as a professor does.
I [41:50]: M-hm.
S [41:50]: And a professor has encountered a lot more, like, training, and a lot more experience to actually grade someone else's work. So, they're a bit more meticulous.
I [42:00]: M-hm.
S [42:01]: Um. And I mean, a professor is teaching a subject so that another person learns the subject. I'm, I'm learning it, so I—I'm not as meticulous as someone else, because I'm also trying to grasp the concepts.
I [42:19]: Well, so with that—you're learning the material, you're trying to be more meticulous. Do you get the feeling that your peers are doing the same when they grade your videos?
S [42:30]: Yeah. I mean... for the videos that we've had, like, the grades that we have up till now, I feel like... yeah, the comments have been, um, like—also in the grading, you can tell who like took the time to watch your video, and who didn't.
I [42:48]: M-hmm.
S [42:49]: Because you can obviously tell, like this person said that everything was good. That it was... satisfactory. But then the comments have nothing to do with what, like, your video actually included. But then you can see someone who was like, okay, you're very good, like, try to include a little more in the conclusion, or try to include a little bit in this part of the video or that part. So I feel like some people do and some people don't.
I [43:18]: What do you feel about the people who don't?
S [43:21]: It's pretty annoying. Because I mean, their videos are being graded too, and we're all taking this class! Because I mean, if it's not because it's part of your major, it's prerequisite for like a lot of other things. And I mean, I'm on vacation, I'm back at home. I'm taking the time to take this class and try to get a good grade. And if they're—if they don't take the time to like actually grade my videos, which will affect my grade, I mean, it's pretty annoying. [laughs]
I [43:52]: M-hmm
S [43:53]: It's pretty annoying. Because we're all trying to get a good grade in this class. I don't think anyone's taking a class during the summer online just 'cuz.
I [44:01]: Right, so—uh, tell me again why people might be taking this class? in the summer?
S [44:08]: I would say because they're, like me they're on vacation, they're somewhere else, and they're not on-campus so they need to like start taking more credits. And that way, like, they can get ahead in some other classes. Or, I don't know, what other... or maybe they're, like, in a summer camp, or doing co-op, and they also want to like take advantage of like the time and start taking more credits.
I [44:35]: Sure.
S [44:37]: Um... but yeah, I can't think of a reason why a person—if it's not because they want to make use of their time, I can't think of another reason why a person would take an online course. [laughs]
I [44:50]: That's very fair. Um, getting toward the end of the hour, I only have a couple more questions I'd like to talk about. This has been—this has been actually a rather exciting interview session.
S [45:02]: [laughs]
I [45:04]: So... your strictness in grading relative to that of the instructors. Do you find that that's changing as you get further along in the course?
S [45:15]: Um... in a way, yes. Because when you do the lab calibration videos, like, you realize whether you're grading higher than you should, basically. Um... but then, I dunno, I feel like I try. I try to think of how the professor would think how to grade it. Um. But in some way or another, I feel like sometimes I—I'm not able to completely fill in that, like, I'm not completely able to like think like a professor does, because I'm also learning, so...
S [45:53]: I would say it is changing, but not fully.
I [45:56]: And, is it changing in a particular direction, one way or the other?
S [46:03]: Yeah, I would say it's changing toward trying to be like a professor. Trying to grade like a professor.
I [46:10]: And what sort of grades does a professor get, if you had to sum it up very broadly?
I [46:22]: So when you're trying to be more like a professor, you're saying very clearly that you're trying to adopt the attitude of a professor and think more like a professor in giving the grades. Uh, but at the end of the day, you're giving a set of grades, and then you're comparing them to a set of expert grades. You—you seem like—you seem to believe that you are less strict than the professor. Does that mean you tend to give higher grades?
S [46:52]: Yeah... again, I feel like I do that unconsciously. Yeah, I feel like I try to like think like the professors in some way, when I see that lab calibration videos when I'm a little bit off, I try to think like okay, maybe I'm giving it too high grade, or something. Um, but at the same time, deep inside, I'm like... this person seems to know what they're doing. Um... and I'm not a professor, to like actually be able to grade too hard on

someone. Um... so, yeah, I feel like there's two sides of me. One that's like, okay, you're not grading as like the calibration videos are telling you to, but at the same time you understand the person, so, I guess... yeah. It's unconsciously.
I [47:46]: So it sounds to me like you're saying that, um... you need to achieve a certain level of expertise before you can... justify giving someone a lower grade.
S [47:58]: Yes. Yeah.
I [48:01]: Does that feel right?
S [48:01]: That—definitely, yes.
I [48:03]: And is that something you felt at the beginning of the semester? When you first started to...?
S [48:07]: Well, I always felt like... honestly, like, one student grading another student. Yeah, it helps you learn in some way, but at the same time it's like, I'm not an expert at this topic. I don't know if I should be grading another person's work. And actually like affecting their grade. It feels strange because I'm not an expert on the topic so... I'm not sure.
I [48:35]: And does it still feel strange?
S [48:37]: It does, it does. I mean, I feel like in school, they also made me grade some other people's work, um... and I felt that way, and now that that they got to college, I actually—I was a little surprised on the fact that we had to grade other people's labs. But, um... I guess I understand that we're learning from it. But it still feels strange.
I [49:00]: It feels strange—does it feel bad? Would you rather it be some other way?
S [49:06]: I mean... I guess I understand that there's like a lot of videos and like, it might be hard to grade that many videos for like, faster than the TAs. But I... I don't know. I'm not really sure. I'd say it... it's not necessarily bad, but it is a little awkward. [laughs]
I [49:29]: Uh... a little awkward... just I'd like to talk about that a little more. Why does it feel awkward? In what way does it feel awkward?
S [49:41]: I guess it feels awkward in the sense like, again, some people take it seriously, other people don't. Like I know a TA or a professor, they would always take it seriously, because it's part of what they're trying to teach us students. Like I actually take the time to read the comments of my videos. Some other people just grade everything satisfactory, and [unintelligible] like that. So it's awkward in the sense that like, yeah I know I'm taking it seriously, but other people who are grading me don't take it seriously. So it's a little annoying to like, have to deal with that.
I [50:20]: M-hmm. I can certainly undertand that. Um... well, I think—I have one last follow-up question. So when you were going through your evaluation, uh, you mentioned that videos—that the people producing these videos were supposed to explain things from scratch. As if you're explaining something to somebody who knew nothing. Can you talk about that a little bit?
S [50:47]: Yeah! I mean, the way technically, like, I approach these videos is like, okay, we're supposed to like give an introduction, as if, like the viewer doesn't know what the experiment is. Um, they're not—they don't—they might have a basic idea of physics, but not really something, um, too deep—so like, I've had videos where like I start it, and it's... it starts from, this is my video, this is the data I got, where like I would assume—what if I'm a person who doesn't know what the video is about, or what you're trying to explain? Um, what—yeah, what if randomly I run into the video online where I'm trying to find an explanation for physics, um, and I find a video that starts off like that. Like, that's not helpful at all.
S [51:42]: Um, also becasue I feel like, um, by trying to explain to someone else, you learn more about it. Um, in the sense that like, if you're not able to explain it to someone else, you realize that you don't really understand it fully either. So, if... you should always try to, like, think of explaining from scratch. So that, like, also you benefit from it. Like, you're not also helping someone else to understand it, but you're benefiting, because you're understanding better whatever it is you're learning.
I [52:19]: And do you think you've benefited from explaining to your fellow students about things?
S [52:24]: Yeah! Definitely. Um, I feel like, by trying—by writing down the script, like that's the way I've done it, I would write down what I'm saying in the video, I've had to stop and think, like, wait a minute, why I am I saying this? Why is this important? And how deep should I explain this? Um, also um... my brother, he's also studying engineering. He's not at Georgia Tech, but um I've talked to him about some subjects sometimes, like some of the physics problems, and I'll be like, wait a minute, I don't know how to explain this, and when I'm able to like fully explain a problem to him, um, without him having to say things, like oh hey, I feel like I understand it. So, it's a pretty satisfying feeling when you do it. Yeah, I've definitely felt like I—um—I've learned more by having to explain things.
I [53:23]: Very good. So is there any other—anything else about your experience in the course or your experiment—experience in this interview that you'd like to talk about?
S [53:35]: No! I think that's it. [laughs]
I [53:37]: Alright, thank you very much. Okay, so I still need two things from you. I need your media release form—email that to me as soon as you can. And also the payment form. Uh, so I can't—I can't just give you money through here, so the way we do this is if you're a GT employee and have have an employee ID and you get a paycheck from Georgia Tech, you can give me your GTID and I will add twenty dollars to your next paycheck. The other option is that you can uh, put your address and we will mail you a check.
S [54:10]: Okay!
I [54:11]: Okay. Please send those two forms, and please also send me—I didn't check to see if you sent your rubric to me yet. Uh, but please do that.
S [54:22]: No, I haven't.
I [54:23]: Okay. So please send all those together as soon as you get a chance. And thank you so much! This was a very good interview I think.
S [54:32]: Awesome! Thank you.
I [54:35]: Thank you.

F.4 Student 4, Summer 2016

Interviewer 01]: We are online now. Okay, well thank you so much for agreeing to participate in, uh, this interview.
Student 10]: No problem.
I [11]: Before we begin, I've sent you a consent form that I'd like you to sign, along with a media release form. So the consent form tells you the nature of the study, and there's no known risk to you for participating in this study. The media release form does something else. So, the media release form, if you choose to sign it, gives me permission to share the anonymized recordings of this interview with other researchers, for research purposes only. So you don't need to sign the media release form to participate in this study, I'll still be able to use your video for my own research, um... but if you do choose to participate in the study, I will need you to sign that consent form.
S 49]: Okay. Um... so should I print—should I have this printed out, or...?
I [55]: You can, uh, an electronic signature is fine. Are you on a Mac?
S [01:02]: No, I'm on a PC.
I [01:03]: Okay, well you should still—if you open up that .pdf, you should still be able to type in your name, or add a text box or something.
S [01:09]: Okay.
I [01:10]: And I will need you to do that before the experiment, so let's do that now.
S [01:16]: So, it's the form that says consent document?
I [01:20]: That's correct.
S [01:22]: Okay.
S [01:42]: I'm not exactly sure... should I screenshare, or...?
I [01:48]: Uh, you could annotate it? Uh, by typing your name on the line, and then email it back to me.
S [01:55]: Yeah, I dunno why but it's not letting my type anything.
I [01:59]: Uh, what program are you using to open it?
I [02:02]: Um... it's just says... um... iCloud, and then... I'm not sure what program... one second... it's the e-pdf. An internet pdf.
I [02:29]: Uh, it's just a flat—a flat pdf. If you open it up in, um, I guess Windows Picture Viewer? Whatever the name the Windows program is.
S [02:44]: Okay.
I [02:44]: And you should be able to annotate it.
S [02:47]: Okay.
I [02:47]: We'll come up with a more streamlined way in the future.
S [02:58]: One second.
S [03:37]: Can I just do it in a paint program...
I [03:43]: Yes, absolutely, you can even just screenshot it and do it that way.
S [06:02]: Okay, I attached it.
I [06:03]: Excellent, very good. Let me just see if it's come in.
S [06:10]: Okay.
I [06:12]: Very good. Excellent.
S [06:19]: Okay.
I [06:20]: Okay. So let's go ahead and begin. So the way interview is going to work—it'll have a few, uh two main parts. We'll begin with a think-aloud interview while you're evaluating the video that I sent you in that previous email. And then we'll have a more free-flowing interview afterward where I ask followup questions and we have some discussions with your experiences. Um, so I'd like to begin now by walking you through a think-aloud problem for practice.
S [06:53]: Okay.
I [06:54]: The whole interview will last about an hour, I'm recording this interview over your webcam. And while you're doing your video evaluation, I will ask you to share your screen so that I can see what you're doing when you evaluate. Would you like to begin?
S [07:09]: Yeah.
I [07:10]: Excellent. Okay. So, uh, first I'd like to warm up with a practice question, so that we can get used to the way think-aloud process works.
S [07:20]: Mm-hmm.
I [07:22]: Please get something to write with if you don't have it already.
S [07:26]: I do. I'm ready.
I [07:29]: Cool. I'm going to ask you a physic squestion, and I would like you to narrate to me the thoughts going through your head as you solve it. Uh, your thoughts don't have to be coherent, you don't have to express them in complete sentences. During your problem-solving, I will not ask you any other questions except if you're being quiet, from time to time I might prompt you to keep talking.
S [07:52]: Okay.
I [07:52]: Okay. Uh, let's begin. So the physics—the sample question is a two-part estimation question.
S [08:00]: Okay.
I [08:01]: What is the volume of the room that you're in right now, and if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for the room to fill with water?
S [08:14]: Okay... so, um... to estimate the volcume of the room I'm in, I would probably... measure the width of it, going from the side [gestures] and then from the side, and the height. And I would just basically multiply that in order to find the volume of my room. And... if... I were to turn on a faucet...
S [08:46]: ...and try to figure out... how much water, or how much time is needed to fill up the entire space... I would need to know basically first... the rate at which the water is coming, and then... by knowing that, and knowing how—the mo—the volume of my room, I would try to calculate it by doing that. Like... somehow. Doing that. In order to solve it.
I [09:23]: Okay.
S [09:24]: Yeah.
I [09:25]: Uh, okay, thank you, that was very simple, and that's the way that a think-aloud process works. Um... and you did very well. So now let's move on to the sample video evaluation. So I've sent you a lin—a copy of a video for you to watch, uh, along with a copy of the usual uh, lab report rubric that you've been using in class.
S [09:48]: Okay.
I [09:48]: And correct me if I'm wrong, I should have I sent you—you know, it looks like I didn't sent you—
S [09:57]: Yeah it doesn't look like it's got the rubric.
I [09:58]: Let me do that right now.
I [10:11]: Okay, I've sent it.
S [10:18]: Okay. Yeah, I've got it.

I [10:19]: Okay, so you will see the expert evaluation after you're done with your evaluation. Uh, your evaluation obviously will not affect your grade in the course, but please do try to evaluate it as if you were doing a practice or calibration assignment.
S [10:36]: Okay.
I [10:36]: While you're watching and evaluating the video, I would like you to narrate to me what's going through your head, just like last time.
S [10:42]: Okay.
I [10:42]: When you're watching the video, when you're actually paying attention to the video, it's actually rather difficult to narrate out loud what you're thinking, you don't have to talk too much while you're watching the video. Um, so with that said, would you like to begin?
S [10:57]: Okay, yeah!
I [10:59]: When you're doing this, when you're watching the video, please share your screen with me. There's a toolbar on the side, the second icon is the screenshare icon, but before you click it—when you do click it I will be able to see everything on your screen—
S [11:12]: Yeah.
I [11:12]: —so make sure there's no, uh, personal information on the screen.
S [11:15]: Okay. [inaudible]
S [17:59]: Okay, so I'm going to pull up the rubric now. And is it okay to like go back and forth with the video?
I [18:18]: Yes of course. And when you're filling out that rubric, could you please type that up in a separate text document and send that to me later?
S [18:18]: Oh, okay. Yeah. So...
S [18:21]: I'm just going to do that after like, I... I like say whatever, and then...
I [18:28]: M-hmm.
S [18:30]: I'll just write it down. So, organizational, I felt like it was very organized, however I don't remember seeing a statement of result. I'm gonna look back.
S [18:46]: From here... Oh, there...
S [19:30]: Okay, so she does, actually. And... lack of introduction, there was introduction... conclusion... Um, it's, for me I felt it was logical in structure, and just the few transitions sometimes she cut what she was saying, earlier than... um, than like allowing her to like finish the sentence, so that just caught me off guard a little. But overall I would give it a good. Um... I'm going to go type that.
S [20:13]: Should I just type it in the email below?
I [20:16]: Yes, that'll be simplest.
S [20:20]: Okay.
S [21:48]: And content model. She disc—she does discuss the principles, um... but as far as, uh, the connection she makes to the model, I felt like she could have better emphasized on that. Um...
S [23:01]: What I think she could have done better is, when whe was explaining the energy principle, she could have talked—or like shown a better visual of it. For example, like, for the energy principle, just even stating delta E equals delta of gravitational potential energy, spring potential energy, and kinetic energy. Those changes. She could have done that. And she did that later, in like explaining her what does it mean, or... one of the what if questions. But, I felt like, um... previously when I looked through, uh, the videos that other students had done throughout the course, they would always include it in the beginning. So it would be easier to just look at that and then, um, view the video and... the programs.
S [25:42]: She definitely did explain the... possible errors that could have caused her data to be a little bit off from the computational model. And... the computational model from the actual video. Um... I don't believe she talked about how she, um... the specific, like, initializations in her visuals, like, um... besides the spring's, uh, stiffness and how she calculated all those, maybe she could have also included... like, when she, she could have actually just shown the program itself, and given like the stuff that her—I guess, kind of unneccessary, but it's better to have, like... the g equals nine point eight or, et cetera, like that. She...
S [26:54]: And like, how she initialized it, how, um, just the code part, basically, uh, the how she got, like the [inaudible], the arrows, or what they mean. But she did a good job in explaining how she got the csv file. Um... But besides that, she definitely did a great job in explaining all the different, like, when the—when she—uh...
S [27:28]: ...ran the program, and, uh, saw the actual spring model and explained how what each arrow means, and um, how she possibly could have had errors, um, between hers and the video itself. So that was good.
S [28:01]: But not all data was—say, used to initialize the model are clearly identified. So, with that, I would give... overall I would still give a good.
S [30:31]: Content overall. Um. I have a question. So... for... the, this question, I think this was what does it mean part, was it the same question that we had?
I [30:49]: Um, yes, I believe it was—yeah, there wasn't a change between that semester and yours.
S [30:59]: Okay, so maybe... even... for the what if, I definitely got her explanation, but um, the what does it mean I remember it has to do more of a why we chose, uh, the... x-direction analysis, rather than the y, and that had to do with the y being, um... the components of it being the mass and k, so... the spring itself, and then x being gravity and the length. I believe. And, it didn't seem that she had any explanation for how you calculate the x direction and the y direction for both. She basically just discussed the how—how thermal energy could have been an issue.
S [32:42]: Yeah, she said it might have been due to plotting, but she could have further explained that by describing the different formulas used in order to find the x and y directions, which is a key think I feel like she missed out on. So, there are—there is good discussion of both questions, but, um, as far as "contains no physics errors" she should have included her explanation of the different formulas for those. So, I would give that a fair. Or... mmm. Compared to, uh, the calibrations I used to do on the Webassign, it's a little bit, um, stricter, I feel like. So I would actually say that I'd give it a poor.
S [33:58]: [unintelligible]
S [36:14]: Uh production delivery. Overall her powerpoint was very well done, and there wasn't really any issues with the audio besides, um, sometimes she would cut into [unintelligible], I think she did it once here. I feel like she could have had better transitions than that. But besides that, the audio and video were really good, so...
S [36:54]: I'll just give that a good.
S [37:28]: Um, let me see how long this was, I think it was... just on time.
S [37:51]: Okay, so I'm done.
I [37:54]: Okay, excellent. If you would please send that to me, I will now send you the corresponding expert grades for that same video.
S [38:01]: Okay.
I [38:02]: And if you would please compare the two, or do what you would normally do when you see the expert grades.
S [38:07]: Yeah.
S [38:12]: Okay. So, results in intro, very good, overview... should have spent more time covering principles... where did the length come from? good calculation of k, some sloppiness in language, without a very clear connection... okay.
S [38:33]: Confusion with... missed x/y spring/pendulum behavior. Some audio gets cut off, slides a bit dense with calculations. Okay. Now let me compare that to mine.
S [38:57]: Oh I did... the difference was—the main difference from mine was the content model. [muttering] Not really the content model, but definitely the content overall was very off. But, okay.
I [39:20]: Okay.
S [39:21]: Should I stop screenshare.
I [39:25]: Yes. please stop screenshare, we'll move on to the follow-up part of the interview.
S [39:31]: Okay.
I [39:31]: Excellent, So, uh, now we're just going to have a more informal talk about your exp—the experience you just had and the experiences you've been having in the course.
S [39:42]: Okay.
I [39:43]: I noticed a few things happening when you were evaluating that video. Um... you watched it once all the way through, then you didn't remember if you'd seen something so you went back and sought around for it. And you actually did that a few times. Is that your normal way of doing it?
S [39:58]: Yeah. Um, I usually just uh go through it once completely, sometimes even if like, you can make the video go faster, so I'll just make it go faster and for one time, just look at it overall, and then I'll go back when I'm doing the rubric. I'll see specific spots where I thought, oh, I think it was different, and then I will like confirm that.
I [40:27]: And have you always done that? Thinking back to the very first lab.
S [40:31]: Um, yes, most of the time yeah.
I [40:34]: And when you're going through the rubric, you also went point-by-point for each item. Is that, again, something you're comfortable doing? Is that what seems natural for you?
S [40:45]: Yeah I actually use the web—on Webassign, they would give us an even, like, detailed rubric, and it would have like, everything from what good should be to what very good should be to what excellent is, for each, like content model, overall production, and whatnot. So I would look at that some—for at least the first one, and then for the second as I... do the next video, I like, keep that in mind and then I like sometimes I even like compare the first video to the second, and I'm like... if I gave this person this, should I... like, when I'm comparing it, did I think that one was better, so that's like kind of what I did.
I [41:30]: When you did that, did you also compare with your answers to previous labs?
S [41:39]: What do you mean?
I [41:40]: So, if you're on lab 2, you might evaluate, you know practice 1, practice 2, calibration 1, and then you're not sure what grade to give so you'll compare that, you know, practice 2 with calibration 1 so it's consistent. Would you do that in previous labs as well, or just within a single lab?
S [41:57]: Yeah. Yes. I would definitely—I would look at what the teacher, the professor had to say, and then how he graded the—I remember he'd like, specifically whichever one is like th ebest video out of the calibration models, I'd look at that and his answers, or his replies to that, and I look at the video that I'm currently looking at, and I'll be like, okay, well, the—compared to that one, this does not have the result or this one does not include the pricniples, or whatnot.
S [42:33]: Sorry [phone rings]
I [42:35]: And so it seems to me like, uh, you're doing that because, um, consistency is very important to you. Would you say that's—why do you spend that amount of effort making sure your grades compare with your other videos?
S [42:53]: Um, I feel like... if one of them, if I, I judge the first one the best because I started out with that one, so, um, I feel like my answers to those are the most accurate of how I feel, like, the professor would grade it, and how I would grade it, so oftentimes I'd think—I'm not, I also don't want to like, give someone a worse grade, but like... just going through it and being like, fair, poor whatever, so I feel like if someone else is grading my work, then I would definitely want to grade them, equally and fairly. So...
I [43:37]: So do you think, when your peers are grading your video, do you feel as if they're putting as much effort into the grades as you do?
S [43:45]: Yeah.
I [43:48]: Why do you feel that way?
S [43:50]: Um... I feel like, I know my peers that I had for my group and for the group problems that'd we do, and they were pretty um... I feel like, just like talking to them I really feel like they would definitely work—like, I don't spend like ten minutes on video, I spend like probably seven or five, but I definitely do go one by one and clip through it and I feel like they would do the same. Just based on working with them, so... um, and it's part of our grade, so I would—I wouldn't want to like, just take that lightly, so.
I [44:33]: Okay. I also noticed you said that, uh, when you were thinking—there was one point in this evaluation where you seemed to talk yourself down from a fair to a poor. Um, and you mentioned that you thought the webassign videos, or the ratings of the webassign videos were a little bit stricter than the ratings you give. Can you tell me about that?
S [44:53]: Yeah. Um... I'd like, usually whenever I submitted and then I look at the review, it's usually my—grading, I'll say good, good, good, and then the answer—like, the professor's grade would be like, Fair, Poor, Fair. Or like one, one below what I say. So I—if I thought it was fair, I thought—because she didn't include, I felt like a crucial part of it, so I was like, okay, think fair, and then so it'd be poor. I was like, yeah.
I [45:33]: But in general, do you think that the... um, why do you think that is? Why do you think the instructors tend to give lower ratings, uh, than you do?
S [45:47]: Um, well, I mean, besides like, just the rubric, they have a better—because they're the ones who created the assignment, they probably have a better answer in mind that they want. So maybe if you don't live up to it, I guess? Because I've never seen any of the calibration or any of them give anything above a very good. I mean I've never seen

an excellent. I mean, even like the ones i gave excellents to, they would give good—I don't know why, exactly, maybe it's because they're trying to... it would, for them, their idea of excellent would be something different than mine, so I guess I just grade easier, so... I felt like I would compare it to that.

I [46:41]: As, so you guess you grade easier. How does that—so you might have an attitude, you might have made a judgment about a video, but when it comes time to put in a grade, if you're thinking that you grade easier than the experts, how does that effect the grade that you give?

S [47:01]: Sometimes, like I did with this one, I would—if I gave a fair, I would maybe give a poor. But um, most of the times, I also think of it in the way that they're also grading my labs, so... I tend—I also definitely compare how I would—what I produced, to what they produced. And so, if I think they were better than mine, I... if I would give myself, like, probably because—I mean, obviously you should not be biased, but if I'm giving myself a fair, even if I feel like a poor, and I give them the same, I would not give them poor either, so, it's just—in that sense, yes.

I [47:47]: So why—do you want to give your peers higher grades?

S [47:52]: Sometimes. Sometimes, yeah.

I [47:54]: Why is that?

S [47:56]: Because they're also grading mine! So, yeah.

I [48:01]: And you want them to behave the same way as you do?

S [48:05]: Yeah... and definitely like if I feel like you generally didn't put in the effort, I'm not going to just give you a good, or give you a fair. Like, in that sense, I'm not completely—I felt like I'm never completely off from what the professors would grade it. If, for like something say, just that, even that um... video in which I was looking at the what does it mean question, I gave her a poor maybe if I was like, knowing my grade was also dependent on theirs, I would have given a fair, and like—it's up to me, like, I'm not sure how, it just...

I [48:48]: Does it seem fair to you that you are in some part in charge of other people's grades in this course?

S [48:56]: Um... yes, because they're also in charge of mine. So... in that sense.

I [49:05]: Would you rather have... uh, had grades directly from the instructors, with no peer evaluation?

S [49:13]: Um... no. I feel like um, even besides just grading it, even though, um, they would have done a probably better job and sometimes when I look at the lab feedback, it'll say something like "good job"! And it doesn't further explain it, like you've given me an 80

I [50:07]: In the sense that they might notice things and deduct points that a student wouldn't notice?

S [50:13]: Yeah. Yeah, m-hmm. And then also, when I'm like, after doing the feedbacks, even if they were to do the professors, even if the professors were to do the grading besides the students, they should definitely do the calibration, I feel like, because um, I felt like thoughout my videos, they have progressively gotten better because of doing the lab feedback, and like knowing what they want to see exactly. Because of—when I like, I was just looking one time at the last video I did, the motion lab, and I had the link to my lab one, and i just randomly looked at it—and comparing it, there's a complete difference in how much I've like put effort into it.

I [51:02]: Oh! Can you tell me a little bit more about that difference?

S [51:03]: Okay. Um, so, basically my lab one I just read—I did not have like, clips into... say, the introduction, and then, so I like organized it by like Powerpoint, and then I had like introduction, principles, the tracker analysis, the part—the... uh, program itself, and explaining each of what I did. But like, the first one I just ran through it like, once, and like, I um... self—wasted—once, and then I just put that and sent that in. And it was kind of sloppy, you could like, see I like stopped sometimes, and then I would just move on. It was not transitioned fully. And I completely forgot the what if, and the what does it mean, too. I just thought those were just extra questions that if you just wanted to say, you could. But I didn't even look at the rubric before doing that. And then I like, after doing the calibrations, I was like, oh, okay, I need to include this and that. So yeah.

I [52:12]: Very good. So, with that, when you first came into the course and you learned that there was going to be laboratory exercises with peer evaluation, what was your—what were your expectations of that?

S [52:29]: Honestly, I didn't know exactly how it would be graded—I was like, first I thought of it, oh they just want our opinions on our grades, and they—the TA herself or himself would grade it actually, so—the first lab, I would say was a complete, whatever, like mess. I graded—I definitely paid attention to the calibrations and everything, but altogether I was—'till the point that they like actually gave the grade, I didn't even know who was actually grading. It wasn't explained.

I [53:08]: And was that distressing? Did that make you feel—how did that make you feel?

S [53:15]: Yeah, kind of, because professor Greco said that that he, um... put in all the grades before the first test, so everything was there except for the lab, so I was just like, I'm not sure like, honestly I felt like maybe I'd get a fifty on the first one, compared to like the rubric, and the calibration I had done. So, I definitely like going in I was just nervous that it would just bring it down.

I [53:45]: Tell me, so, nervous that it would bring your overall course grade down?

S [53:50]: Yeah.

I [53:51]: So what were you—so in terms of getting a good course grade, when you came into the course, what did you expect you would have to do to get a good course grade?

S [54:01]: Um, well, altogether I felt like there was so many resources provided that and like... so many things that we need to do at like a certain date, that if I just do those on time, and um, just use the resources that are provided, I would—um, at least get a B. In the class. And like, I'd aim for an A, but... I would just be, yeah.

I [54:30]: Okay. Um, I have only a couple more questions. Uh, I noticed you said earlier that you had never seen an excellent when you're going through the expert evaluations. The experts never seemed to have given an excellent.

S [54:45]: Yeah.

I [54:45]: And... uh, so you—why do you think that is?

S [54:53]: Um, I felt like—I feel like just because an excellent would be like, a completely—there's no changes necessary to be made on the video. But um, in their sense. But maybe like I view my definition of excellent, would be different, I guess. Like, if they do all—if they follow all the things, and if there's like a little bit of a error, I don't think that an ex—they wouldn't—they should get an excellent for that, because like, yeah. But, I felt like, for excellent they mean that there is no changes that can be done. So maybe that's why they haven't given one.

I [55:36]: And you said, the experts might have had—or the instructors might have had some expectations, and maybe the students wouldn't live up to it.

S [55:44]: Yeah.

I [55:44]: You used the phrase "live up to it".

S [55:44]: Yeah.

I [55:45]: What sort of expectations do you think the instructors had for you?

S [55:49]: Um... expectations, uh, well like, how... maybe they have already like, not, I don't think they have, but like, maybe they already have a video in mind, or have already produced one, and like from that they look over to the student—it might, compare to it. As grade.

S [56:21]: You're lagging [glitch].

S [58:11]: Okay. Yeah I think the video lagged.

I [58:12]: Yeah, sorry, we're back online now. We only have a couple more minutes.

I [58:18]: So, we were—we were talking about what you felt the instructor expectations for y'all were.

S [58:25]: M-hmm. I just said that, I know that they didn't have their own video, but maybe they had like... better like—like, the rubric is broad. In the sense that like, it's the same rubric for each one, but in their mind they might have like, this principle should be explained, and how they talk about this question, they should have this... a better definition, or explanation. And if you don't have those, then you don't have like, you don't get the certain like, good, excellent, very good.

I [59:09]: So something like a—like a secret rubric, or a more detailed rubric...

S [59:15]: A secret or like in their head, they should have this. They know that—why don't you have this? Because we're like the ones that are like learning this topic, the physics right now, so we might not have think oh this is necessary to have. Because like even looking back at my lab one or my lab two, I'm like, why didn't I include this principle? Like obviously after taking the test and already studying for this, I should have done that. Like, why didn't I use that. So maybe like, because they have the prior knowledge of what to include and how to analyze something and how to make the connections between the model and the physics behind it, they grade harshly in that sense maybe.

I [60:04]: So, "harshly", does that—I'd like to talk about that word a little bit.

S [60:10]: [laughs] okay.

I [60:11]: Do you think when the instructors give a harsh grade or a low grade, do you think they have a—negative attitude, do you think that's mean?

S [60:22]: No! No I don't. I just like—I mean it's like... obviously they know what an excellent should look like in their minds. And they know—they've been teaching this course for years, so... they know what they want to see in the lab and like, the analysis of it. So... in that sense maybe that's why how I grade it is different because they have like, an expert's idea of what to do and stuff so... yeah.

I [61:02]: Very good. Um, so if you were to—knowing what you know now, and with the experiences you've had, if you were to go back to lab one and view the same videos again that you were originally assigned, how would your grades or attitudes change?

S [61:19]: Sometimes, if I didn't know the like, entire like, explanation of the physics behind a model, but I like knew I should include this principle or something, and I would sometimes second-guess, oh, if I thought that they had an error in their explanation, I'd be like, am I correct? But I wouldn't like go and open my book and be like, okay, this is wrong, definitely. So going back if I like saw something and I knew for sure they should have included this, like I did in the lab—the lab we saw, um, I would do that and I would know exactly for like the content model, and the content discussion, I felt like I would have a clearer picture on what to grade. And like, how—like, my... my score would be more accurate.

I [62:17]: Okay. Um... very good, thank you very much, those are all the follow-up questions I have, unless there's something else you'd like to talk about.

S [62:25]: Oh, no. Um, so for the other two documents, the media and the payment, should I—do I need to do anything for those?

I [62:40]: Yes, you do need to turn... if you would like to turn in the media release form, then please fill that out and email it to me.

S [62:43]: Um sorry, I didn't know exactly what you were saying for the media release form. Do I need...?

I [62:53]: No, so the consent form says you've given consent to be interviewed by me, you've consented to me saving this recording and using it for my researching and sharing it with my lab group only.

S [63:08]: Yeah. Okay.

I [63:08]: What the media release form does is, um, at some point in the future I might be giving a talk or talking with researchers outside of Georgia Tech, and I might want to show some clips from the interviews that I've been giving.

S [63:18]: Okay.

I [63:20]: Now, that's not—you don't have to agree to that to participate in this.

S [63:24]: Okay, okay.

I [63:26]: So the media release form you'll either release to me or allow me to do that, or not, either way it's perfectly fine. The other form is the payment form, which I imagine you want to fill out because you want your twenty dollars.

S [63:36]: Yeah.

I [63:36]: I can't give you the money through here, so on the payment form, you can either indicate that you're a GT employee and we can add twenty dollars more to your next paycheck, or you can give a mailing address and we'll mail you a paper check.

S [63:50]: Okay.

I [63:51]: Uh, so please fill those both out electronically, and send them to me as soon as possible.

S [63:56]: Okay. Thank you!

I [63:58]: And thank you again, I think this was a very good interview.

S [64:02]: Thank you, have a nice day.

I [64:04]: You, too.

F.5 Student 5, Summer 2016

I [02]: Now. Okay. Uh, so thank you so much for participating in this interview. Uh, before we begin I have sent you a consent form I'd like you to sign, along with a media release form. So the consent form tells you the nature of the study, and tells you that there's no known risk to you for participating in the study. The media release form is optional. If you choose to sign it, uh, the media release form gives me permission to share clips of this recording with other researchers for research purposes only. You don't need to sign the media release form to participate in this study, but if you do decide to participate you do need to sign the consent form.
S 40]: Okay.
I [44]: And you can sign it—you can either type your name on the line, or if you have a tablet you can sign your name to it.
S 51]: Okay. Gotcha. Um... lemme see if I can edit this.
S [01:12]: I'll have to export this to the... to a doc, and then I'll be able to type on it real quick.
I [01:18]: Okay sure. You can also... are you on a Windows computer or a Mac?
S [01:24]: Windows.
I [01:26]: There should be... there should be a text box or annotation tool. You can type your name in there as an annotation. Whatever works—easiest for you.
S [01:36]: Okay. I'm using the um... Acrobat, but not the pro version.
I [01:46]: Oh, okay.
S [02:54]: Let me see if I can just upload it to something and then type it in on that.
I [02:54]: Alrighty.
I [03:08]: I wish there was a simpler way to do this but this was the best we could manage.
S [03:12]: Sure.
I [03:12]: I'll see if I can get an interactive form.
I [04:15]: And if you have any questions about the content of the form itself, please ask me.
S [04:20]: Alright, will do.
S [05:30]: Uh, I'm almost there, sorry.
I [05:31]: No worries.
S [06:31]: [inaudible] pdf. Okay, and...
S [08:03]: ...attachements. And that's sent off.
I [08:11]: Okay.
S [08:12]: Gimme a second.
I [08:15]: No worries.
S [08:15]: Okay, now it's sent.
I [08:17]: Okay, excellent, let me just see when that comes in. Sorry that took so long, uh, I can definitely come up with a better system next time.Um...
I [08:34]: Unfortunately I do have to administer consent before you—we begin the experiment.
S [08:40]: Yeah that's fine.
I [08:41]: Okay. I'll just wait for it to come in.
I [09:16]: And you sent it to my gatech email address? You just responded to the email chain?
S [09:22]: Yeah, I just responded. It should have gone, hopefully.
I [09:25]: Uh... still updating, let me just...
I [09:45]: Uh, nothing yet. It was just a pdf that you sent as an attachment to the email?
S [09:51]: Yeah, no text. It'll—the thing it just the normal consent, and then it has edited at the end.
I [09:57]: Okay, um... still haven't received it. Why don't you send it to—oops...
I [10:17]: How should we do this? Um.
S [10:24]: Can I send you stuff on this? [presumably over Hangouts]
I [10:24]: You should be able to, yes. If you have a link to something, or if on the sidebar on the left, you can bring up a chat window and then if you can paste links to things in that windows.
S [10:37]: Okay, let me... let's see, where should I put this...
[skipping ahead, technical issues resolved]
I [12:30]: Okay, so let's get back to the study. So the way this study will work is that uh... the first thing we'll do is I'll walk you through a think-aloud practice physics problems, just to get used to the way think-aloud studies work. Uh, then I'll ask you to evaluate a sample lab video report like the video lab reports which you've been doing this semester. And then, finally we'll end with a more freeform sort of interview where we discuss followup question that I have, and we'll also talk about your experience in this study and in your course in the previous semester. The whole interview will last about an hour, usually three-quarters of an hour to an hour. I'll be recording this interview—I'm already recording it right now.
I [13:16]: And for part of the interview, when you're evaluating the video, I'd also like to record the content of your screen so I can see what you're looking at when you evaluate the video.
S [13:25]: Okay.
I [13:26]: So would you like to begin?
S [13:28]: Yes.
I [13:29]: Sure. So first, I'd like to warm up with a practice problem so we can get used to the way think-aloud—get used to the way the think-aloud process works. So get something to write with if you don't have it already.
S [14:14]: Alright, ready to go.
I [14:14]: Perfect. So, I'm going to ask you a physics question, and uh, I'd like you to narrate to me the thoughts coming through your head as you solve it. Your thoughts don't have to be coherent, you don't even have to express them in complete sentences. Uh... during your problem solving, I will not ask any further question, except if you're silent for a little while, I might just prompt you to "keep talking."
S [14:43]: Okay.
I [14:44]: So let's go ahead and start. Here's your question. This is a two-part estimation question. What is the volume of the room you're sitting in right now, and if you were to turn on an ordinary kitchen faucet, about how long would it take for this room to fill completely with water?
S [15:07]: Okay, so first question being, what's the volume of the room that I'm sitting in. Um... I guess the way you could start to think about it is, well I'm looking around, and getting out some type of measuring device. Like a... some tape—a tape measure, and starting from one dimension, go to the opposite end, and then to see what that measurement is, write it down. And then going from uh, perpendicular wall, uh finding that dimension and going to the other end. And then taking the meter stick—or, the measuring stick, and going straight up. And that way you get um, a measurement of the volume. And then as far as the second question, how long it would take for a faucet—first, you kinda have to know, um, how fast the water's flowing, and how much.
S [15:57]: So you can kind of test, maybe with a cup that you have—that you know the dimensions of, fill that up and see how long—One, how long that takes, and how much water that is. And from the volume that you've calculated from the room, you can kind of estimate the time it would take and how much water that is. By converting—by how—uh, fast I guess, you can say... meters—or, milliliters per second. You can estimate it that way, and say how much it is for the room itself.
I [16:28]: Okay, very good. So that's how—that's how the think-aloud process works. Um, so now let's move onto the actual video evaluation. So in the email I sent you earlier today, I attached a—file... movie file of a sample video lab report from a previous semester.
S [16:47]: Okay.
I [16:48]: Um... so... I have a little script here... so you will see the expert responses to this video after you're done with your evaluation, much like during the practice and calibration phase of your assignments in the course this semester. So your evaluation of this video will not affect your grade in this course, because of course the course is already over, but please do try to evaluate it as if you were doing a practice or calibration assignment. So while you are watching and evaluating this video, I'd like you to do it in the normal way that you would in this past semester. And I'd like you to narrate the thoughts going through your head as you're evaluating the video.
S [17:30]: Okay.
I [17:31]: And I know that when you are watching and paying attention to a video it's hard to also narrate your thoughts about it while you're watching it, so I don't expect you to say very much while you are watching the video. But when you're evaluating or when you're thinking something to write down, please let me know what you think.
I [17:48]: Okay, so please—and while you're doing this, I would like you to... I would like to share your screen with me so I can record your screen. Before you do that, um, I will be able to see your whole screen when you're doing that, so please close or hide anything that you don't want me to see.
S [18:04]: Gotcha.
I [18:05]: Okay. And on the toolbar on the left, there's a screen share icon, a little green icon with an arrow.
S [18:13]: So, sharing?
I [18:15]: Yes, perfect.
I [18:24]: And so please open the video, and please open a text document or something, and type out—when you're ready to evaluate, please type out your ratings and comments. I also included a copy of the rubric...
S [18:36]: [plays video, tries to stop it] ahhh... stop...
S [18:39]: Okay. Let's...
I [18:52]: And in the email, I also for reference sent a picture of the rubric that you've used this semester.
S [19:00]: Okay, yes. I saw that one. Are we ready to go?
I [19:06]: Uh, I'm ready if you are.
S [19:08]: Okay.
S [19:18]: Seems like there's a lot of text on here. Um, at least she puts them in three different steps, so that's good. Kind of giving a little introduction.
S [20:48]: It's good how they use cutouts of screenshots of their line, instead of the whole entire code. It's easier to read that way. As well as the snapshots of different things that are written down.
S [21:18]: Explains all the different backgrounds, um... relations that are needed to get all of the values needed for the code.
S [21:54]: It'd be nice to see the... um, Tracker video as she tracks it in realtime. Though this isn't necessary since she gave her data [inaudible].
S [23:20]: Nice to show the graphs really large, and then also put it in screenshots so you can get the important data she's talking about. It's nice that she has, um, the principles there at the top so you know what's going on.
S [24:12]: [inaudible] the dissipating force which means she added something in that kinda takes away from the total force.
S [25:13]: Let's see. So, um... looking at it, she has a good beginning middle and end. Explains the conclusion at the end, although it's kind of over time. But um... but if it was still at the same timeframe to be what it needed within five minutes, but... I think overall was explained pretty well. And the um, video itself was done really well. Alright.
S [25:41]: And... that's it.
I [25:45]: Oh, so, um, please go through and assign a rating to each of the rubric items, as if you were filling out a proper calibration rubric.
S [25:55]: Okay. I will do that.
I [25:57]: And then after that I'll show you what the instructor rating for that video was.
S [26:03]: Okay. Rubric... [inaudible]
I [26:34]: You might be able to right-click and open it ina new... tab?
S [26:36]: Um...
I [26:37]: Oh no no, you've got it.
S [26:39]: Okay.
S [26:49]: Alright that was terrible.
I [27:03]: Let me email you just the rubric itself. I think it might be automatically compressed to save space.
S [27:09]: Yeah.
I [27:28]: Okay, sent.
S [27:30]: Yep. Let's see.
S [27:36]: Alright. Good. Alright. So let me bring up my thing... so as far as organizational um... structure, like I said she had a pretty good uh beginning, middle, and end. In terms of setting up the, um... the lab. Spoke really well. Um, had transitions. Kind of, um, it kind of flowed naturally in terms of how the, um, lab went. In terms of what she needed to calculate. But as far as if someone didn't know what this project was about, it might have been a little jarring to see why you go from one setup to another. Especially not showing the video itself, except for a screenshot. But um, overall, I would give that a very good.

S [28:21]: On the um... organizational structure. And then for the content models, let's see... identified models... the model issues were relevant were relevant, all the calculations were something that were needed. Physical ideas, so calculating things like the initial uh, length, the starting length of the spring and all of those spring measurements were in there, and talked about. As well as bringing it back to fundamental principles. And also the energy principle. So I'd probably give that a, um, very good as well.

S [29:01]: Okay. And then for content prediction... It was kind of uh put at the ends, at some of the big discussions, especi—and there wasn't a section on the discussion questions that are usually given at the end, as far as a slide, but it was kind of talked about throughout the presentation. And kind of just understanding what's going on with the lab and why things are heppening, but just a couple things that are kind of... uh, not put on a slide individually in terms of discussion, so I'd probably give that a good.

S [29:38]: And then... let's see... then content overall. Um, I believe this is probably in the very good category. Um, it seems like everything was in there in terms of what needed to be explained. Um, just a couple things that were kind of explained but not to the fullest extent. Um, so overall the questions were answered but there could have been a couple more content—like adding in the what if and what does it mean. In terms of its own slide, but it's kind of explained a little bit...

S [30:10]: And then as far as the production delivery, I would say that's excellent. There weren't any error—well actually, let me step back. I would say very good to excellent. Um... everything from audio was great, the video itself was great, there was just that one jump cut where that was content that was kinda cut out, or it just wasn't a smooth transition. So... um... it's very, it's basically almost excellent. Uh, it just needs I guess it needed to be looked at one more time to kinda get that last thing out. But there was nothign that was distracting. Everything was on there. There could have been at some points a little bit less text, but it's not overbearing, and you can still follow along what's going on.

S [30:57]: And I believe that is it.

I [31:01]: Okay. So if that's your evaluation, I'm sending you the instructor evaluation—uh, we're not done with the screenshare quite yet. I would like to... we'll continue the think-aloud as you look at the instructor evaluation which I have just sent.

S [31:20]: Alright so I see very good on the intro, that's the same grade I gave her. Um, result, like I said results in the intro. Content models says good, I said very good. Should spend more time covering the principles, connections between the principles and motion. So, yeah, I guess... it couldv'e gone further, is what this is saying.

S [31:44]: Um... and then... we have for the CP is good, I had good as well, and where does data come from? And good calculation... discussion, k. I guess she does skip over where that L nought—she does explain the calculations, but I guess there could have been more of where does this number come from. Especially, there are some things that are kind of given, but she doesn't explain those, so it's kind of within the lab but... things that could have been explained in the video so the average observer could know.

S [32:17]: For the overall content we have—sloppiness of language, "mass of the period"... uh... E over E confusion, not very clear connection to motion, unclear discussion of heat transfer. And missed x y pendulum behavior. I guess there was that one thing where she didn't really explain why there was a difference between the x and y in terms of the swing, um, mass of period—these are kind of, I guess things I would say myself as a student, but maybe not the correct uh terminology for physics. So I guess I missed that part. And then unclear connection to the motion—she does explain a lot about I guess the physics of it, but maybe not the motion itself in terms of being kind of in an oscillating pattern. And then for the production value, says audio gets cut off, I think that was when the jump cut happens. Slides are too dense... in calculation. Headlong narration... [inaudible] Okay so like I said I was thinking about this, she does have a lot of text-heavy parts. Um... and it's kind of a lot on some pages, but I think to me, it seems it's pretty clear what she's saying as long as you're listening intently and kind of just marking as you go along.

I [33:37]: Okay. Thank you very much. So now, now you can stop screensharing. Thank you for doing the think-aloud portion of the interview. Now we can move into a more conventional interview portion. Um... so to begin with, I have a few questions I'd like to ask that I noted um, when you were doing your evaluation. So, when you were watching the video for the first time, you were also typing comments along to the side. Is that normally how you do these things?

S [34:06]: Yeah, I'll probably note it on a piece of paper. It wouldn't be on the actual screen, it would be something that's very small, just in case it's something that very much jumps out at me. But it's not for every video, it's only um, if I feel the need, it's not something like I set up to do. Um, but... if I needed to I would write something down about it, and it would just be like little scribbles like "note this", "note that", or um... well actually, usually um it wouldn't be on a piece of paper, it would be within the actual lab report, where you do the evaluations.

I [34:43]: Right.

S [34:43]: I would just start typing into that, and then, um, come back so that way when I watch the video I don't forget and especially watching a lot of people's videos, just putting it in there, like I noticed from this section needs to go into that as a comment, and then come back later.

I [34:59]: And you watched the video—you just played it from beginning to end, you didn't rewind, you didn't skip around. Is that also normally how you did it in the semester.

S [35:08]: Yeah it's definitely just one playthrough, unless I definitely missed something. But usually I would just kind of listen to it, one go-through, and then note what I'm looking for, and then just kind of watch it as I needed it to be, without playing it all.

I [35:26]: And when you say, unless you definitely missed something, how would you know if you definitely missed something?

S [35:31]: Just like a lapse in concentration, I'm just assuming. Especially if there's something very noticeable that's wrong with the video. So if there's like—

I [35:41]: Like what sort of thing?

S [35:42]: Yeah, huge audio issue, like it's really quiet, or um... there's a lot of stuff going on. Like if someone has multiple screens open while they're trying to do the video, and they're switching between screens, I kind of notice that more. And then I might lose some of the words that they're saying. So, um, it's only with like, I guess a very badly done video. Or something that has an obvious, like, an auido or visual element about it.

I [36:07]: Okay. Um... so for the last rubric item, you said very good/excellent, and you looked like you weren't certain which one you should do. Can you describe to me your thought process there and why you weren't... why you didn't decide to just give one or the other?

S [36:25]: Yeah. So for me, thinking of excellent, it's basically the 10, it's the perfect video, there's no issues at all. So like I said in the evaluation, I did see the jump where it's like audio cuts out a little bit, and it goes directly to something else. And there's just a lot fo information that's kind of on text. So it's more the content of the actual video than... for me the last section is more about video quality. So audio is good up until that one point, and then the video was just very clear. Um, so it's kind of like that one thing is kind of holding it back from like the perfect, excellent—that's what I categorize that as.

I [37:04]: And have you always felt that way throughout the semester, or do you find yourself holding excellent to a higher standard, do you think?

S [37:14]: To me, excellent has always been the unattainable metric. Of... I definitely haven't given it to too many videos on too many different categories, because there's all these things you can kind of count off for. Um... so most of them stop at very good. I don't know if I've... I'm trying to think back if I've given excellents. But if it is, it's usually in the production quality, because that's something that's a little bit easier than knowing ever single thing that goes into the physics that's easy—it's easier to have good—if you have the tools, to have good audio and good video. And to have—

I [37:53]: So you think it has—so why do you think it is easier to have good audio and good video than it is to have good physics?

S [37:59]: Um... it's, it's a more um... as long as you have the right—like I said, as long as you have the right setup. So if you're screengrabbing using like screen-o-matic, that's what I used, and what the girl in this video used. It presents really clear video. Um... and then for the audio situation, if it's recorded off of a good microphone, then it's gonna be good audio. Um, so just if you have the tools, then it's good. If not, then it can hurt the video a little bit.

I [38:30]: Alright. Um, and so do you know... so when you're comparing your ratings of the video to the instructor ratings for this video, on most of the rubric items you had given a higher rating than the instructors. Can we talk about that a little bit? What do you think about that?

S [38:48]: Um, I think it's just because the instructors of course know more than me. Uh, they've been through seeing a lot of different videos on the subject. Or they've done this lab before many times. So they know every little thing that needs to be dissected, which on my part I don't. I know the general things that are like obvious, that need to go in the lab, but there are some things that you need to say this one little thing about it, or there's other things. So there's stuff I feel, um... as a reviewer, it would be better to know, and that way you can kind of deduce it more in line with what the instructor does, which kind of experience, in terms of, the instructor knows all of the things that have to happen in this lab, we may or may not know as much as what needs to go into the lab.

I [39:32]: So you're thinking it's a matter of the instructors knowing what to look for, that's part of what the instructors do.

S [39:41]: Yes.

I [39:42]: So you mention—you said that the instructors know what to look for, they know every little thing that is supposed to be in the video. A while earlier, you said... when you were discussing your response to the overall content, the instructor rating said that there was some sloppiness in the language. And you said that, that makes sense, as a student you don't always use the correct terminology. Uh, can you talk about that a bit?

S [40:05]: Yes. So, um... I basically, I mean being a physics introductory course, there will be some people who are like newer to the subject, it's not like a grad-level or like senior-level, so they may not be as well-versed in the terminology that you maybe use at every instance. And especially once you're kind of recording a video, if you didn't do like a million takes, you might just kinda say something as you're going. So it might not be the, I guess, scholastic level, or like the scholarly like level of saying... that it needs to be said in this way, so that it's a lot clearer, it kinda... people have the tendency to maybe like ramble, or not ramble but kind of say the easier words, of what it means, but it's not the exact definition of what's proper. I guess in the book or in our lectures.

I [40:54]: Can you give me an example of what it might mean to use "easier words" to describe something?

S [40:58]: Um... so... a lot of the time I guess, like saying the "thing" is moving this way, or uh... the "object" is behaving in this manner instead of what kind of object, um, explaining what because of maybe like the shape has some influence on how it's moving. And in terms of motion or describing the motion in a certain way, uh, just different modifiers that bring it in more to the physics realm instead of like, you're telling it to somebody on the street. Uh... so there's kinda that difference between, if you were just like telling someone about your lab versus needing to submit a lab report. So it's I guess report language versus common language, I guess is the word.

I [41:44]: Okay. So, um... let me think for a moment here... So returning to something you said just a moment before, before we got onto this topic, you said that the instructors know what to look for, they know about every little thing, and that's why they give... uh... that's why they gave lower grades than you did here. Do you think that means the instructors are looking for errors to take points off, or do you think the instructors are looking for um... they're looking for features and then adding points as they find the features?

S [42:25]: Um... I think possibly starting with kind of like statement saying what are the errors in this... um... that seems to be just from how the comments are, there is still—I mean it's a little bit of both. You see, especially if it's a lower score like a fair or whatever is below that, uh, you can definitely—it's just errors, and then it might say one thing good. Uh, so it's kind of looking for what's wrong with the report. Um, is how I see it. In terms of like when I was going through the semester and seeing what I would grade versus what the instructor would grade on it, it seemed this was more critical in terms of what's going wrong with it. There might be a couple comments especially if it's on, uh, very good or good or maybe excellent, but mostly what's going bad about the video or something that could be added to the video to make it better.

I [43:19]: Okay, and so you say you think the instructors, us, are more critical than you. Do you think they're more critical than the other students? Do you think the same rule applies to the other students in the class?

S [43:31]: Um, it depends I guess on the students' maybe level of enthusiasm for the course, or how invested—especially, say if a student is going to do physics for a while longer, physics two or physics three or something like that, they may be more invested to learn the different ins and outs of the lab reports. So what needs to be said, what doesn't need to be said, so they may be—as or more critical than the instructor, but as far as myself, um, this is kinda just a class that needed to be taken. It wasn't something that I followed up with—I'm not going to take—I didn't take the physics two course, I'm done with that. Uh, I graduated, so... it's just kind of learning, like, what's good enough it kind of like maybe the normal student. They explained the big steps. You know, like the [unintelligible] like how does this connect to that thing. A normal student may or may not look at. Cuz like that's what I saw in a lot of the comments from the instructor, was how do you connect this motion to that principle, uh which a normal student may or may not have picked up on, they might just say, well they talked about this, this, and this, so it's good.

I [44:46]: So when you got grades back for your own video, when you got the grades and the comments back for your own videos, did you feel like you were being graded by enthusiastic students, or more normal students? What do you think about that?

S [44:59]: Um... I think more um... normal's I guess, like an average student would kind of give you like, "he got all the big points in there." So I guess that's just the realm of students that were like, you got the big points in there, it wasn't completely terribly, I could understand what was going on, so that way they were like, that's a good enough grade right there. Uh, there might have been a couple of students who were like, you missed this, this, and this, and it lies more in like what the instructors were looking for. So it's kind of like a minority of students that are... really looking for all the de—like they are in line with what the instructor thinks in terms of what they're looking for the video. But uh, the majority being students that are looking at the big points, and making sure you didn't mess up. Connect those big points with big discussion.

I [45:49]: I like this distinction you're drawing between maybe the majority and minority of students, the really enthusiastic students, and the normal students. It seems like you are suggesting that the majority of the students are... looking uh... maybe looking for reasons to give each other a good grade. Would you think that's a fair thing to say?

S [46:11]: Yeah I mean I don't think... I don't think it's that enthusiastic students are trying to bring students down, they're just more aware and are looking more—I don't know if "more invested" is the word, but more into how the instructor's grading, and they're kinda emulating that. Whereas other students are like, well you got the general idea, I can understand where you're coming from, especially since they're doing the same thing at the same time. They can understand where you're going with the project even if it's not evident completely in the video, so they're more likely to say, alright, you got the point down, I understand, and then give a slightly higher grade maybe than an instructor or a student that's really critical on the video.

I [46:54]: And so do you think people in general—not people in general, do you think the students in, your classmates... were they inclined to give each other good grades? Do people want to give good grades to their peers?

S [47:08]: Yeah I would, for myself I can't really speak for anyone else, but I'm not looking to destroy someone's grade. I'm looking to say you got these general points, you explained—you connected things together, so I'm more inclined—as long as you have the basics of what the lab entails, and then what's going off the rubric. I look at what's going on in the excellent category on the right, and saying okay they hit this this and this, so I'll give them a slightly higher grade. Um, and kind of looking at what's on the uh, zero, on that scale of the mark—of the marginal. I'm looking more at the advanced category of like, hopefully they hit all these things, and I can grade it based on that. But kind of, uh, yeah just looking at... what they got in the advanced, and then kind of scaling it towards that.

I [48:01]: So I'm going to ask a slightly provocative question—why don't you want to destroy other people's grades?

S [48:07]: Well I guess, personally at this point I don't care. Um... about other people's grades, because this is my last class in terms of undergraduate. So... uh, and then especially on an online class, I don't know any of these people. So I don't have any personal vendetta or anything about them. It's just... might as well just give them... because were all kind of in this together. In terms of, need to get this pass so that... I don't take joy in like people not passing a class. Um, so kind of just as long as they did the normal things they needed to do in the lab, it just gives them—It's not, I'm not trying to say like if they didn't do something sufficient I'd still give them a higher grade, but if they kind fo just meet where I was in that lab, or explain better, that's kind of where I set it up. Like, where I was in terms of instruction, and then if they did more than that, then make it a little bit higher on that section. If they um... explained a thing really thoroughly, connected different dots together, then they get a slightly higher grade on that portion.

I [49:16]: And if—and so you say this online course, you don't know anybody else, it's maybe not the most competitive environment. Were some of the other courses you've taken at your time at Tech, were they more competitive?

S [49:28]: Um, yeah, definitely um... big electric classes, because especially if there was one that was on a curve. Um, it's just kind of—it feels more competitive, it may not be actually more competitive, but people don't want to—like people don't want to study together, something like that. Like I need to get a good grade, and maybe keep my knowledge with my friends. It might feel more competitive, people aren't as likely to group up or talk with other people, um... but I guess, especially with physics, it gets dispelled if you're in your recitation, um, you kind of are more likely to see what other people are doing, and maybe form groups. It's kind of a counterbalance in terms of um, online and offline, so offline it's kind of—especially since people are just gone, um, it wasn't like in-session, people weren't on campus. Uh, it's kind of different, where it's—you're just in it together, especially with the hangouts with the, uh, weekly assignments, the GPS ones. You have to figure out the assignment with someone else. So they might as well be together on it, and actually get it done because you don't want to spend like thirty hours trying to get it done.

I [50:43]: So, in that sense, as you're—would you say that your experience in this online course compares favorably or unfavorably with your experience in other kinds of courses?

S [50:55]: I would say this one, definitely more favorably. Um, in terms of, at least comparing this one to another lecture-style class, large lecture-style class, because especially since lectures were kind of videos given, and you could watch them at your own pace, and then go back and watch them again, so the need for taking notes was a lot less, and maintaining notes throughout the whole entire semester was a lot less of a burden. Um... the collaborative nature of the GPS was good. Um... uh, the only thing is timing, in terms of those were—we had ours at night, so some days it would kind of be uh, not beneficial to have them at night when we want to go do something else, or you have some type of meeting or something during the day. Um... but overall I liked this experience because it made the large, um, lecture hall seem a lot smaller, because it's just you just kind of seeing the videos from the instructor, or you're in small groups. So it's kind of like if physics was only recitation-sized. Which inevitably makes it a lot easier to kind of learn from other people and talk to other people, because you're not having to talk with a hundred or more people.

I [52:07]: Okay, very cool. So just a couple more questions. So when you first joined the course, and you say that you would be doing peer evaluation exercises, what were your general expectations for that?

S [52:18]: Um... I'm not going to say negative, but I guess neutral. It seems like it needed to have—it needs to happen somewhat with an online course, to see what the feedback it. It was kind of a lot, especially with the grading the labs. Having to do a calibration and then looking at the peer reviews from that current lab. So... that the [inaudible] so you know what's going on, and how the system worked, but having to do it every week, over for every lab, was kind of a lot, I guess. To kind of go through the calibrations and then go through the actual evaluations of people. Um, and then beyond that, I guess the only other ones were the midterm the evaluation and the final CIOS, which were normal, so that wasn't anything more than what I was expecting. But I guess... the only thing I would say about the evaluations this semester was, do the calibration either just once, or not as many per week, so you could calibrate maybe one or two, likely. A good one and a bad one. So you can kind of see a range of like where you need to score in, instead of I think it was four or so, four or five, so I guess just less on that, um, so you can spend more time on the evaluation itself. Because say it's, if it's due on tuesday night, and you haven't done it until it's tuesday morning you have to do it, you can kind of get burned out doing it all at one section, so I guess ideally it would break it down a bit. So just making it easier, especially with an online class it's not going to be—you're not going anywhere, you're just going to your laptop, so people probably tend to um, especially asking this in Piazza when it's the day of, an assignment's due, and everyone's asking questions about assignments that are due in like a couple hours, so you're more likely to kind of wait till the last minute until it needs to be done.

S [54:19]: So making it... not less work, I mean, less work on them to like actually do a good job of grading. So I think if there's a lot less things to grade, I might put more effort into each individual one instead of spreading it out between, that would be like ten evaluations basically.

I [54:37]: So do you think um... so... do you think that the calibration and practice videos are less valuable as the semester goes on, is that what I'm hearing from you?

S [54:46]: Um, yeah. Cuz you kind of understand what the instructor's looking for once you look at their comments. It's still nice to have at least one, like I said one or two each lab, so you can understand what's going on for that specific lab, in terms of what instructors are looking for, but not I guess as many.

I [55:06]: Okay.

S [55:07]: And if it was not required but you could at least look at a reference, maybe you—it would give—I'm not sure, it would depend on the person, people may just not look at them at all, then. Um... at least have the reference, but just less of them to actually have to do.

I [55:23]: So you say some people might not look at them at all, some people mioght not do whatever the assignment is. As the course happened, as it really happened, do you think the grades you got from your peers was trustworthy?

S [55:39]: Um... overall I think so. Um maybe towards, maybe the last one, they were kinda like, it was the last lab, let me just give him whatever grade. It might have just been straight down the middle, or maybe all good, maybe all bad, it doesn't depend on the person. But probably—maybe that lab, the second to last, it was really hard, and they were like I don't know what's going on, maybe. They just, they just kind of stairstepped down on what grades to give. And put some random comment on it. But overall I think people where kind of just—they were doing the lab as it needed to be done. Or, the evaluation. And... the best of what they thought happened. Their grading on that. Overall.

I [56:26]: Okay, very good, so those are all—um, those are all the substantial questions I have. I have one minor one, you used earlier the acronym GPS assignments—I'm not familiar with that.

S [56:38]: Um, it may not have been GPS. It was like Group Problem Solving, I guess? I guess so—

I [56:43]: Okay yeah, the hangouts where you would do problem solving with a group of peers.

I [56:50]: And secondly, unless there's anything else you'd like to add, or if you have any questions about this process, I think that was a successful interview.

S [56:58]: Okay. I don't have any questions at this time.

I [57:02]: Excellent, thank you for your time. I need to more things from you, I need the media release form—well, I need you to decide whether you want to sign the media release form, and if you do, please send it to me. The second thing is the payment form. I can't just, you know, give you twenty bucks here. If you fill out the payment form, if you are—well you're graduated, so you wouldn't be, but if you were a GT employee and you got a paycheck from Tech, I could add twenty bucks to you paycheck, or I can put in your address and I can mail you a paper check.

S [57:34]: So I had been working on campus, so I guess my last day was yesterday... um... I should still be on direct deposit.

I [57:43]: Let's not risk it, why don't you just put in your address and we'll mail you a check.

S [57:48]: Okay. Okay, that's fine.

I [57:51]: Okay, so please send those two forms to me, and thank you very much for your time.

S [57:57]: Okay, thank you.

F.6 Student A, Summer 2017 (Pre)

Interviewer [01]: Excellent. Okay, uh so what we're going to be doing here today, uh, this is a multi-part interview, it's going to take about an hour total, uh we'll begin with a warm-up question so that I can walk you through the sort of interview procedure we're going to be using. It's called a "think-aloud" procedure.

I [22]: And during a think-aloud procedure, I'm going to ask you to discuss a topic or solve a problem but, crucially, while you're doing it, I want you to narrate to me the thoughts that you're having as you do it. I won't be—I won't give you many prompts, I

won't be an active interviewer. Mostly what I'll do is if you're quiet for a little bit too long I'll just remind you to keep talking.
Student 46]: Okay.
I [47]: Um, so that will be the first part of the interview. A warm-up, think-aloud problem. After that, the lab report video that I emailed you earlier today, I would like you to watch it all the way through, and then evaluate it as if you were evaluating it for the lab assignment that you've just done.
S [01:07]: Yeah.
I [01:07]: And while you're watching the video, you can just sort of watch it silently, I know it's hard to think aloud when you're watching something. But when you're filling out your rubric, uh, and formulating your responses, I would like you to think aloud, the decisions you're making.
I [01:25]: After that, there's a much shorter video of a ball rolling across the screen, I'm sure you're very familiar with that sort of work by now. And I would like you to construct an explanation of the physics underlying that system, much like you would if you were making your own lab video.
I [01:40]: And then I am going to ask you some questions about that explanation, and we'll discuss various topics related to physics and to peer evaluation and to your lab exercises. So that'll be the last part of the interview. And that will be that, after that you can sign your consent—sorry, your media release form, and then we can schedule your next session which will be at the end of the semester, so it's going to be a little while now. The goal here is to, uh, capture an interaction with you at the beginning of the semester and at the end so we can see what sort of changes our course effect in people.
I [02:16]: Okay, so. The warm up question for the think-aloud... okay, uh, I'm going to ask you a physics question and I'd like you to narrate to me the thoughts going through your head as you solve it. Your thoughts don't have to be coherent, they don't even really have to be in complete sentences. All I'm going to do is prompt you to keep talking if you're a little too quiet.
I [02:39]: Okay, so, let's begin.
I [02:40]: This is a two-part estimation question. What is the volume of the room that you're sitting in right now, and if you were to open an ordinary kitchen faucet in that room, about how long would it take for that room to fill up completely with water?
S [03:02]: Okay! Well... um, it's kind of—sitting—it's kind of an L-shaped room, I'm in the, like, there's like a... kinda like a reverse L-shape, like there's another corner... and then there's the front door over there, and a door over here. So it's kind of a big room, I would say that this section that I'm sitting in is about... let's see... eight feet long. Another eight feet wide, and eight feet tall.
S [03:31]: So I'm kind of sitting in the cube section, in this part, and then there's like a little divider but it's not a separate room, and I'd say that section of there, it's a lot bigger... it's probably about twelve feet long, in this direction [gestures], maybe twenty to twenty-five feet long that way, and then twelve feet long on that same wall again.
S [04:01]: So let me write this down, it's about eight by eight by eight, and then this would be... [writing] twelve, twelve, let's say in the middle of twenty and twenty five is like twenty-two point five, and it's eight foot tall throughout. So.
S [04:14]: The section I'm in—it's kinda like—got a little diagram here, it's like eight by eight [shows drawing to camera] and then twelve, twelve, twenty-two point five. So eight cubed would be the volume of this section, and then eight cubed is five hundred twelve. So five hundred twelve—this is in feet, so five hundred twelve feet cubed, and then twelve by twenty-two point five by eight would be [manipulating computer] twenty-one sixty feet cubed. So the total volume of this room would be twenty one sixty plus five twelve, is twenty-six seventy-two feet cubed.
S [04:59]: So that would be, um, [mumbling, holding figure to camera.]
I [05:04]: Okay, excellent.
S [05:04]: There's some other stuff I wrote up there. I dunno why. So twenty six seventy-two feet cubed, and I guess if a kitchen faucet was turned on, it would kinda depend on how fast it was going, like, if it's a drip versus like a [gestures] you know, but assume that we turn it on at full speed, and just it's going...
S [05:22]: Then I would say maybe it would fill about... hmm... thinking it would fill about... I would say maybe... two feet cubed per minute? There would be some sort of constant per minute that it would fill the room at, so let's just set it at two feet—maybe not per minute...
S [05:43]: I would say it would be four feet cubed per minute if it was going really fast, so... four minute—four feet cubed, per minute... so to find how many minutes it would be, you would just divide, uh, twenty-six seventy two divided by four. To get this, [mumbling, using computer] That's about six hundred sixty-eight minutes to fill this whole room, and six sixty eight divided by sixty is eleven hours. So that would be eleven point one three three... hours. So, yeah, [inaudible] Kinda just did it about [mumbling numbers, holding figures to screen].
S [06:30]: And I mean, it depends on the speed of the faucet, it could be five per minute, but you could just substitute that for any rate that it's kinda going, I was kinda doing a guesstimate of what a kitchen faucet would do.
I [06:43]: Excellent! And that's exactly what this problem was about. Um. I'm always tempted to keep asking followup questions, there's a lot going on there, but, uh, that's a whole study unto itself, and that was just the warmup. And so I think you've got a very good understanding of what sort of, uh, is expected of you during think-aloud studies.
I [07:01]: Okay so let's move on, thank you very much, let's move on to the next part of the interview, which is watching and evaluating the video that I emailed just a moment ago. And for this part, uh, I would like you to switch your hangout so that it records the screen, so I can see what's up on your screen when I'm doing that. Before you do that, make sure there's no sensitive information visible on your desktop. Because that's going to get recorded.
S [07:27]: So I just need the video open? I don't need the rubric open at the same time?
I [07:31]: Uh, you can keep the rubric open if you like. I just want to see your whole desktop. Do it as if you were evaluating a video for your lab, you know, if—referring back to the rubric, typing, pausing. Try to do it as naturally as possible.
S [07:46]: Okay
I [07:46]: Yeah.
S [07:46]: Is it okay if I write down notes on the [gestures to notebook]
I [07:48]: Yes. Yes. Um, if you—anything that you write down—that's actually great, most people type notes up. If your'e writing it down, if you could somehow send me a scan or picture of any documents you produced, uh, during the interview afterward, I'd like that.
S [08:03]: Yeah of course.
I [08:05]: Thank you.
S [08:07]: Okay. Alright. So, I just screen sharing...
S [08:13]: Uh, yeah. Okay.
S [08:20]: Alright, is this good?
I [08:22]: Uh, yes, that's perfect.
S [08:24]: Alright.
S [14:29]: Alright, so I go to the rubric now.
I [14:31]: Yes, I'd like you to evaluate this video and think aloud while you're doing it. Well, it was... I did think it was pretty good. I mean, it was kinda—we haven't really covered the energy principle yet, so some of those things, it was—she talked a little fast, and since I don't really understand I was like wow, she's really going. But for someone who had like known this principle, studied it, I'm sure that it would have been more fine. And in terms of her like—production [mumbling]
S [15:02]: But everything else looked really good. Her visuals were very good, very organized. Some of the slides were a little bit busy, and that could be overwhelming to like an average viewer, but otherwise I think she had a lot of information. She explained herself very well.
I [15:17]: M-hmm.
S [15:17]: And one of my favorite things was like the introduction slide? Like she had a really good intro. I would say that was an excellent intro because she clearly stated the fundamental principle, what results she'd be getting, and then what steps she'd do the problem and get the results, so I think that'll probably be an excellent.
S [15:36]: I guess I can go over... checkmark [manipulating rubric image] [mumbling]
S [15:46]: But I do think that would be excellent.
I [15:51]: I think it's the button to the left of the one that's highlighted.
S [15:55]: This one?
I [15:55]: Nope, over on the left. The very far left.
S [15:58]: Oh, okay.
I [16:00]: I think it's that button right there that you need. I think.
I [16:05]: Mmmm.
I [16:07]: You know, you can actually just write down on a piece of paper your rating and comment for each item.
S [16:16]: But yes. I do think it was a very well-structured video, and she did a good job of laying it out at the beginning, and what steps... so I think that was excellent. And she was also... pretty good at uh identifying what sort of principles were going on, like how it was represented in the code, like she put the code next to every step of the way, like, I think on the energy graph—not the energy model—but one of the graphs she put the relevant lines of code in the corner and I thought that was a really good way to like connect what the fundamental principles and like discuss the principles while that... so I thought that was a really good way of showing how... what you're doing in the code, and what you're doing in python, how it would be relevant to the motion that you actually see in the lab and what you've plotted.
S [17:08]: And I think—yeah she did a good job with discussing the physics ideas, and she identified the models overall, the energy principle, and she had all the formulas in there as well, so I don't know what else to say. That was an excellent. [mumbling] I'll rate that as an excellent, the models that she used... [writing]
S [17:33]: Okay. And for the content, she did um... one thing that I would have liked to have seen in the video was the actual motion of the spring in real life. I think she just had a screenshot of her tracker video, and it didn't actually show the video of the motion of the spring. And I—she showed the motion of the model when it was going, and the vectgors that were associated with that, but she didn't actually show what the model of the spring in real life looked like, or at least I didn't catch it.
S [18:06]: So I think I would have liked to see that, just to see what the initial motion could have compared, just in a qualitative sense. To the actual motion in the code. And... she did a good job discussing how the x-component... in the model was... had a different... sort of matchup with the y-component, since I didn't fully understand what the energy principle was beforehand, and I feel like afterwards I still don't really understand it, so I guess if this was some sort of video trying to help people understand—like completely understand what the energy principle is, I feel like that... wouldn't really be the case, because she did talk a little fast, and did discuss the—I was more focused on the visuals I guess rather than what she was saying, so I guess if someone wanted to have [inaudible] or saying more simple slides, it would be better, because like clear structure...
S [19:03]: Overall it was very good. She did discuss whether the model predicted the motion pretty well. I would say that was very good. [mumbling]
S [19:23]: Alright. And, I wasn't entirely sure what the what if—I didn't have the actual... you know, kinda, what you're supposed to do in the lab. So I know wer'e always given w what if and what does it mean question. And I know I usually label that, the what if, I don't know if that's the right way to do it but she did have some good discussion of the fundamental principles, she had the equations, she had everything.
S [19:52]: And I think that she didn't really... judging on the formulas she used, and what her work—I didn't really see any glaring physics errors, rather than what did—the errors that she discussed, and how her code could be improved, and how the model is less accurate over time, and all the errors that would have been present, she did try to explain with real-life examples of what the limitations of the model was. So I would say that that was also excellent.
S [20:22]: And, as per production delivery, she did have very good vocal quality. She did kinda get hung up on some sentences sometimes, but I mean, that's obviously that all people do, but I would recomment that if you do sort of like get hung up on a sentence in real life, [inaudible] I would just take a second take, I would delete that recording. So kinda doing it in bites that were smaller, and like don't bite off more than you can chew.
S [20:58]: And it was very good very organized, but then again she did talk very fast, and did have some errors in the talking, but otherwise everything looked good. Everything was, you know, pretty good. But I dunno, I'm kinda stuck between good and very good for that, because it was kinda... it wasn't poor lighting, it wasn't poor production quality, like everything was very good it was just I guess... vocal qualities could have been improved. So I'll put very good for that section.
S [21:28]: I mean it's just something you can do with practice. Multiple takes.
S [21:38]: I know it's hard to do that sort of thing but I think otherwise she did pretty good, it was a really good video. And also I think it was exactly the six minutes long, which is inclusive of the [grace period]—I think five minutes plus the one minute grace period, so, if she went any longer than I would have said it was too long, but otherwise

[mumbles] pretty good length, she had a lot of stuff to discuss, so I think it was pretty fair that she took so long but... then again, you can always, there's always some stuff you don't need to discuss in detail, but there is stuff you need to comment on and get really into, but... I guess, that's all I really... think...
I [22:19]: Okay, excellent, so I'm now going to send you the expert ratings for that same video, and I'd like you to read them and then react to them.
S [22:27]: Okay.
I [22:28]: I've emailed them to you just now.
S [22:33]: Alright.
S [22:44]: Okay. So, I don't really know how to zoom in on this a little bit... so you can see it better.
I [22:51]: I mean I know what it is.
S [22:53]: Okay, okay. Alright. So yeah I would agree that she did very good with the overall video in the intro. I'm not really one hundred percent sure what else they wanted from that comment? Like she said it was very good—they said it was very good and that she had her results in the into, I don't know what would stop it from being excellent. I've definitely seen a lot that were much worse than—for the first lab report I personally did not have as good of an intro as I should have. Because I [inaudible] the results there. But she did a lot better than I did, which is why I rated it as excellent. Principles [reading]... motion... I do think that she uh did a good job at pointing out what [inaudible] but she could have related it back to the tracker video more. She had her code [all the steps of the way?]
S [23:52]: But yeah I guess that is valid she should have connected the principles with the real life motion and not just code. And at least to a greater extent, like I understand that. And where did the L zero data come from? That is a valid question. You have to discuss where your data came from.
S [24:08]: I do feel like she might have uh... rushed through the tracker data, because she was basically at time, so she might have skimped on the initialization data, and where her real emotion came from and focus more on the other sort of principles like the... model, and the energy principle, in the code. "Good k calculation"... [muttering] And let's see, "some sloppiness in language, mass of the period". Yeah, I guess if you give it a—I guess on the first watchthrough, it's harder to [inaudible] especially since I haven't studied this yet, I'm not a hundred percent sure what's going on, it would be less easy for me personally—like if I had personally done a lab report on this, I could have probably more easily caught these errors, like "mass of the period" and sloppiness of language, but to the average listener on the first listen, I didn't really catch any of these though. But I'm sure if I did listen to these again and not been as overwhelmed by all the visuals and kind of had a more trained ear, like, known what was coming, known what to expect, I probably would have caught more of these little things.
S [25:22]: Like, discussion of heat transfer... and... I don't know if she missed the x/y spring—yeah that was something, I guess. Like she did mention that there was a difference in the x/y spring behavior in the model, like how the y component—like the y spring behavior was less accurate in the model. But I guess, yeah that was true that I didn't really know what that was, and why it behaved in that way. I missed there could have been a... more thorough discussion of what those x and y components were, and why they would possibly be different, but then she was at time, so she would have had to cut down on other things.
S [26:03]: And I would have, yeah, between very good and good for the production/delivery, and I did notice the point where the audio got cut off, but she didn't finish her sentence so it was very abrupt, into the next sentence. And I do agree that they were a little bit dense, and she had a busy narration pace, and... I guess I, again not a hundred percent versed on this principle, so there's always a way you can express your ideas more, like more succinctly as they say. So yeah I would say that that's a fair estimate. I was going between a good and very good for that. So, eh.
S [26:42]: I don't... I did definitely think that the um organization was excellent, but there was... definitiely areas she could have improved on. But otherwise, I guess... yeah I... these are pretty, seems like most of these are pretty fair, I would say. Like, fair in their judgment, not like fair in the grade.
I [27:02]: Right. Okay, very good, if you could please switch the camera so I can see your face.
S [27:10]: And I'm back.
I [27:10]: Okay excellent so, uh, as a follow-up to that video analysis that you just did, uh I've got a few follow-up questions uh I'd like to talk about with you for a little while.
I [27:20]: Okay, so, uh you mentioned early in your discussion that you haven't covered the energy principle yet, so you didn't understand the discussion of the energy principle in the video probably as well as you should have. Or... am I characterizing that correctly? You said that you haven't covered the energy principle yet, in reaction to the presence of the energy principle in the video.
S [27:43]: I guess, well, we've definitely discussed like the fundamental principles like Newton's second law, but we just haven't put it in the context of oscillation, and all of those curves, so I mean it was pretty overwhelming to just see a bunch of oscillations and like, I don't really know what was happening. But we've definitely covered the fundamental principles, just not in that specific context of like, spring and oscillation—like we've covered the spring force, we've covered Newton's second law, so I know how those work. But it was just not in the same context as what I was expecting.
I [28:14]: And so when you're producing these lab videos, and you have a concept like the energy principle as it relates to oscillation that you need to cover, uh, do you think think that it is—is it your responsibility to make the viewer understand? It is your responsibility to offer an explanation? Do you think the viewer of—sorry the author of this video, should have helped you to understand the energy principle?
S [28:42]: I do think that it would have been better if she would have set it up more. Like, she mentioned Newton's second law, she mentioned the spring principle, but it could have been a better explanation of why this was important in this specific [item?] what we would have been using in this context.
I [28:56]: So did you feel that this video was aimed for you? That you were the intended audience of this video? Who do you—
S [29:07]: I feel like it was aimed more at people who have worked at spring motion more, and understand it better, rather than just an average person.
I [29:17]: Is that the case—can you tell—so you've made your lab one video, uh, already. Could you tell me what sort of audience you were making this video for?
S [29:25]: I would probably make it for an audience that has a background in physics knowledge, like they know what velocity is. I tried to explain what Newton's Second law, and the position formulas as best as I could. And I went through all the variables. But I didn't explain like the basics of oh, what's velocity, oh what is displacement. Because I—I think my audience was aimed toward someone who mostly knew about this. Like they knew about the basic physics of—but I don't think a random person could walk in off the street and see my video and like, immediately be like "oh yeah, that's... that's Newton's second law." Or like, you know what I mean. Just like the average person...
I [30:08]: Yeah the term for that sometimes is the "lay" person, the person who hasn't, uh, had an educational background in that field.
I [30:17]: So... this relates to another question that I had, something I noticed you said earlier. Uh, in discussing the length of the video, she was right at time, she'd passed the five minute goal and then had the full sixty seconds of grace period afterward. Uh. You said something kind of off-the-cuff but I think there's something interesting there. You said that when you're making these videos, there's always stuff that you leave up. Meaning, there's always more stuff you can put it. So I'd like to talk a little bit about what you decide to leave out, and how you decide what's important to keep in and leave out of your video.
S [30:51]: Well some some stuff I would probably leave out, I think I noticed this when I was evaluating someone else's video, I would—they kind of explained how the tracker program worked. So I mean, I don't think that's really necessary. They were like, "I used the calibration stick this way, and I used the axis—" I was like I don't really think that's necessary, in terms of like... and they kind of explained how they used python, I don't think that was, they kind of...
I [31:14]: So why were these not necessary?
S [31:19]: Well, I think that if you're just looking at the motion, like you just need more of the data, rather than how—like, it's important to know how you got the data, like what sort of initializations you needed to use, but in terms of what the program you used? I think that it's—could could use like a different motion analysis, you could get the same results if you did it the same way. So I feel like discussing the specific program, discussing the all of that sort of thing is not as necessary as what you do with the data. And I would rather talk about, like, what my initial conditions are, and how I could get those—use those initial positions to get a result, rather than just explain how I got the initial conditions I guess.
I [32:05]: Okay, and so what sort of—that emphasis on the data and the use of initial conditions regardless of what particular software platform you're using... what, what's your goal there? What do you think the goal of these videos is, in a very broad sense?
S [32:25]: I think that it's sort of using these like, these programs to... model a fundamental principle, and model motions, and to connect the fundamental principle with your data, rather than connect the fundamental principle with the program, so, if that makes sense.
[siren passes]
I [32:48]: So the important part is... um, the stuff that's not specific—that's not really specific your setup.
S [32:57]: Yes, yes, yeah. It's just, how do we use Newton's second law to analyze motion and predict velocity and predict the position, rather than how do you use this program... So I think it's more of like, I always include the relevant calculations, I always try and do an example. I think in my first lab I did an example of one of the time intervals at which I calculated velocity and like took them through the calculations. So they know that it can't just be... like you can do this yourself, but it's easier to use a computer program because it's a lot of iterative calculations. It would take you a long time to do it by hand.
I [33:37]: Ah, so, you can do this yourself, it's almost as if you're... um... instructing someone how to do that same laboratory.
S [33:47]: Yeah, yeah. It's not that—I don't think that's the... I don't know if I would say that's the entire goal, but it's definitely... showing your own results, but also something that's really important in science is people can replicate it. So they need all the tools they—like you give them everything you know so they can replicate it. In terms of like—how, like calculations, like how, not just on a specific program.
I [34:15]: Have you given—have you ever been to a conference or in a classroom where you've had to give like a colloquium-style talk to present your own research findings?
S [34:26]: Uh, yeah, I did that in biology class when we were doing some research on Piedmont Park.
[34:32.28] Instructor: M-hmm.
S [34:35]: So we talked about it in our powerpoint presentation. All the methods we used, that we discussed what results we got, and something that they really tried to emphasized was "oh, what equipment did you use, when did you go out, what were the conditions, what was the humidity" stuff like that. So I think stuff like that is sort of important in terms of where error can come from? As well as... what sort of, like, conditions could lead to these results. But in physics it's a little different because the initial conditions would be like mass, and like position, and what's the initial velocity, and sort of... different from biology, obviously.
I [35:08]: Mm-hmm. But the goal of—well, a goal of replicability, uh is important in... this sort of dissemination—don't let me put words in your mouth.
S [35:23]: No that's okay. But yeah, you can just—you can show the computer, "oh I used Newton's second law", but not a lot of people know, besides people who have studied physics obviously, they don't know how Newton's second law works off the top of their heads. So that the couldn't really do it themselves. But if you show them how the law works, then... they can understand it better, and they would be able to... theoretically do it themselves, if they wanted to.
I [35:46]: Wonderful. So I'd like to shift gears a little bit and talk not so much about the production of videos as your practices when you're evaluating the video. So, I noticed that when you watched the videos, you went straight through without pausing, and then you did all of your, uh, responses at the end. Is that how you would normally do these videos in your evaluation exercises?
S [36:09]: When I was doing the evaluations, I definitely paused it. I paused it to write stuff down, I guess I didn't pause it... I'm not one hundred percent sure why, I do usually pause it and take notes on, like I took notes here, I took notes while she was talking I guess which sort of contributed to my feeling that she was talking very quickly, but... usually I'd like, when I'm writing down a note I pause the video. But yeah I do usually do all of the evaluation at the end. It could be better suited to do the evaluation of the introduction, and then pause it... but... I guess I just [mumble] Sometimes I do go back if I was, like if i, when I was evaluating an item, sometimes I'd go back to the video and make sure I was understand what they said, if one of my notes was unclear.

S [37:02]: But yeah otherwise I do all the evaluation at the end, usually.
I [37:05]: And, uh, you had your own set of evaluations, and you got to see the instructor evaluations, and there were—the two of y'all gave different ratings. Could you talk for a bit about the differences between the instructor ratings and your ratings?
S [37:19]: Like, the differences in uh quality? Or the...
I [37:23]: The actual ratings that you gave, like you gave an excellent for item number one, the instructors gave a very good for item number one and so forth. Is there... can you try to characterize the difference, maybe, between how the experts assign ratings and how you assign ratings to this video?
S [37:41]: Yeah, um... well I guess I'm not one hundred—like I said I'm not a hundred percent sure why they assigned very good. I do tend to write as much as I can in terms of like, oh she didn't do this, she did this, but I guess I just tend to focus on the positives. Whereas they've done it probably hundreds and hundreds of times, and they... I've seen videos that are actually excellent, but I feel like I just haven't experienced as much like, content as they have, so they would know what would separate very good from an excellent, and what would separate good from very good, and fair... but I guess I do tend to focus more on what they did positively, and then consider what they did negatively after that, and how that—I guess like what I'm saying is like everyone starts off well for me, but once they do bad—like once they do inaccurate things, once they have errors, it goes down from there.
I [38:38]: And what—
S [38:38]: I guess they probably just start with a blank slate, I guess? Or maybe a fair. I dunno. I just always give people the benefit of the doubt, I guess, maybe that's... maybe that'll eventually correct itself once I... you know, look at more videos, and...
I [38:56]: Well I would like to talk about that. You said you do tend to look at the positive, you do tend to start off, uh giving the benefit of the doubt. Can we talk about that a little bit more—why do you feel like you do that?
S [39:09]: Well I mean I know I've made these videos, I've—I'm in the process of making my second video, and I know how... it takes a lot of hard work to do it. It takes a lot of you know, you've got to do a lot of things, and—there's a lot going on, basically. And I... and I just know that people are trying their best, and I—I dunno, well... I guess since I'm in this position [laughs] I know I would like people to not... I dunno, just go—like have a negative mindset on me, so... I'd like to have a positive mindset like about others' work, since I also have a positive mindset about my own work.
I [39:47]: And do you think the experts, when they're evaluating these videos, do you think they share that mindset? Do you think they have a positive, or a negative mindset?
S [39:54]: I don't think—I think they try and have an unbiased mindset, and they just try and look at it more... more objectively—like I do look at it in an objective sense, but I also... try and like, like I said, try and give people a benefit of the doubt. I feel like... they're not as... forgiving, and they're more likely to find all of the little errors, and... all of I guess just not... start at such a good rating and then have to go down, they kind of try and start neutral? And then, from—if they do good things, then they get a better rating, and if they do things bad they get a poorer rating. This is what I would guess.
I [40:36]: Okay. Um... and, so why, uh... why do the experts, do you think, do that? What is it that causes them to have that different attitude, uh, than you?
S [40:49]: Well I think that like I said before, like they saw—they've seen so many of these. And they, they know that everyone is trying, everyone is trying—like doing their best, but... but there's also like, you have to have accountability I guess. If you don't include this, then you shouldn't get points for it, and I think that they've been, like, they have more experience in analyzing these videos—And! With the physics principles as well. Like if you've worked your whole life with physics, then you would know if someone said like, the wrong thing—like sloppiness, like you'd notice the sloppliness in language more than someone who was starting out in physics. It did seem to miss a couple words, but if you've been like studying this for a long time then you notice when someone uses a word incorrectly or does a wrong unit somewhere. So, I think experience is also a major part in how they evaluate it.
I [41:45]: Have there been times—when you've been doing your evaluations, have there been times when you gave low ratings, you know poor or fair?
S [41:52]: Yeah, I definitely have. I think there was one... like at—I dunno—well he didn't really have an intro. He just kinda said, he just kinda went—I don't like it when people go straight into it so I don't—if there is no clear—like I don't know what you're about to do, you just kinda jump into it. Like—if—kinda like—I, I do read the... the rubric things, and if [inaudible] statement of result, it lacks that, then I'm like—well if it lacks it, then I'm not going to give you a good grade. Like I would love to give you good ratings but if you don't have the things that the rubric says then I'm going to rate you poorly. And if you—you don't have as much content as you should then I'm not going to give you a good rating just cuz... I want to help. But if you just don't have what you need to have, then... they—then it's just your own fault, [inaudible] you know?
I [42:49]: Do you find that you have a different attitude or—not an attitude, um... different requirements for good ratings and bad ratings, because you consistently say that you want to give a good rating. Does that imply that you don't want to give bad ratings?
S [43:06]: Well... it depends on the video, like I think there's a lot of... I guess something that does kind of bias me is if they don't have good production, if it's really bad then I'm gonna... and if it—like—if it's hard to see something, then you can't really understand it that well, and you can't really... give it a good rating, I mean... but if it's just bad, I guess it's just bad. But I guess it's just sort of a bias that I have. Like if it looks bad then [laughs] I'll be more prone to saying bad things about it, because—I guess this is just true for everything, if you, like if you were writing a research paper and you were sloppy in your grammar and sloppy—even if your results were like amazing, and you cured cancer, if you did it in a [air quotes] "bad way", if you didn't have good like, good context, like if you didn't present your results well, then I guess, you don't have a good production then no one's gonna really care about it, and not gonna rate it favorably.
I [44:08]: Alright. Very good, okay so we are actually, um, we've got about twenty minutes left in our scheduled hour. This has been a very illuminating conversation. There's one more thing that I would like to go over with you, and if you could please switch to your screen again, and open that much shorter second video of the ball rolling.
S [44:32]: Alright.
I [44:33]: Yes, that's the one. So, I would like you to, um, look at this footage, and look at the Tracker footage that comes at the end of it, and I would like you to explain to me the dynamics and the physics of what's going on there, as if you were preparing an explanation for a lab video. And then we'll have a discussion about your explanation.
S [44:55]: Okay.
I [45:02]: And, and—this is a think-aloud section, so please...
S [45:05]: Okay! So... is there anything else after this?
I [45:09]: There is, there's a dead period, and then there's the tracker footage.
S [45:13]: So it looks like they're doing the tracker. It looks like at first, well, they push it with their hand at first, but then it kinda rolls with a constant velocity at first, but then it starts to decelerate—not decelerate, it starts to stop... accelerating, and kinda... the velocity kinda gets... lower. Kinda starts rolling to a stop in the time interval that we looked at. As you can see by the—there's like a curve over here, so you can tell that the velocity value is smaller, and the position is changing... well, it's not rolling as quickly. And I think the first [part of the?] video... I did like how they, um, didn't include the whole points where the hand was touching the ball in their analysis, because then there'd be different—it would just be more complicated to do when there are other forces at play besides just the... I assume the force of friction, of air resistance, and the force of gravity and the normal force that would also go along with that.
S [46:16]: So it looks like we have the force of gravity in our system. It starts at the origin and goes to about point five meters. [inaudible] And, it was, but they cut out the points where the hand touched the ball, and the initial acceleration—they kinda started when the velocity was more constant, but they also included the times when the ball was slowing down, because that was just the [inaudible] it took, but I guess if I was evaluating this as a constant velocity motion, I would not include those extra frames where the ball was kinda slowing down.
I [46:57]: M-hmm.
S [46:58]: But it's not a huge curve, but it's not like... that... but like otherwise if I was gonna evaluate this constant velocity motion I would want more about the... if I were including the frames where it was kinda slowing down, it would be important to include like what was the friction, what sort of surface are we sliding on, as well as the mass of the ball, and sort of what force gravity is putting on it. Because that's all related to the normal force which is related to friction.
I [47:28]: M-hmm.
S [47:28]: But when—if it was just a constant velocity, um, analysis, I would kinda—I would guess—I dunno if there's a specific point where I would cut it off. Um... maybe around there [pointing with cursor to tracker plot] basically around where the ball starts to slow down, or just focus on the constant velocity motion right there. And... if it was moving at a constant velocity, then obviously the net forces on the system would be equal to zero.
S [47:57]: So that would be less important to figure out what the F net was, whether it it was moving at constant velocity because we would know it was equal to zero. And... the motion I would definitely use Newton's second law to—if I was predicting position and velocity of the object, well, velocity would be constant so maybe just the—the velocity—since the final velocity would be the initial velocity, and the F net would be equal to zero, I would probably use the position formula. Newton's first law? Which would be the final position is equal to the initial position plus v average delta t, and v average we could calculate that using... what... kinda like the displacement formula? When you do like, v two minus v one over... I think it's t two minus t one. So that would give us the initial average velocity.
S [48:50]: But we could also get the average velocity from any two points of data. And we would use the initial—you'd put that in code using ball dot, like ball dot p o s, ball dot vel, you'd use that to track—and ball dot vel would be a constant value if this was moving at a constant velocity, obviously, and ball dot p o s, we could use that—we could use the Newton's first law to predict what the position would be in the future, and model that.
S [49:19]: But I think that if we assume that it was at constant velocity in this motion, it would be—it would be kinda like a straight line, and it would move in a straight line. And it would be a little different than what we see here, since there's a little curve. And there would be a little discrepancy between the model if we just chose to use one velocity for the whole thing. Cuz it does curve a little bit. And...
I [49:43]: Yeah. And so if you're—if you had thirty seconds to provide an explanation of this footage, and that tracker line, uh, to one of your labmates, um, what would that explanation sound like? Now that you've had a chance to look at it for a while.
I [50:03]: Alright. So there's a tennis ball rolling along in the left side—let's set this as the positive x-direction, there's a tennis ball rolling along in the positive x-direction, it starts—the the time interval that we're looking at it at starts at the origin, and it rolls to some position about point five meters in the positive x-direction. It starts off... rolling with a constant velocity, and in this case the F net would be equal to zero and the velocity—because the velocity is constant. We could use—specifically use the position formula with an average velocity to find the new position of this object. However, at the end it slows down a little bit, so there are more forces in play. We would need to calculate the net forces that are acting upon this object, including friction, the normal force, and gravity. And... using that net force, we would use Newton's second law to predict the future velocity by inputting that into the Fnet. And then from that velocity, um, in that time interval, we can use the position formula like we did before to determine what position the ball would be rolling at. And this graph would... explain what the position values were, and we could get the velocity from there. In the experiment.
I [51:15]: Okay, very good. So you mentioned, um, when you were explaining your observations, you mentioned a lot about how you would model that same motion computationally. Um, and actually wasn't explicitly one of the things I asked you to do, so why did you—uh, include that in your explanation?
S [51:38]: Um, I guess, um, I'm just so used to doing the labs where we have to model it computationally, that I guess I just subconsciously connect the tracker data—like if I'm using the tracker program, I guess I just subconsciously connect it with modeling in python.
I [51:55]: M-hmm. And I also noticed that there was, uh the actual act of putting the position points in Tracker, you didn't—you didn't have a discussion there... I would like to have that discussion now. So why... um, if the ball is, you know, it's a discrete object, uh with spatial extent. But yet it's represented by a single point in the center. Uh, I was wondering—could you talk to me about why we're justified in, in doing that sort of simplification? What that means about motion.
S [52:31]: Well, since like the—if it was rolling at constant velocity, the f net would be equal to zero.
I [52:36]: M-hmm.
S [52:36]: And, just, to do a point mass—doing a point mass would be more accurate when the f net is equal to zero, because we don't really need to care about what the radius

of the ball is, because we're not calculating all the forces associated with it. But we have to make sure that we press when we're doing this center of mass, this point mass, [for the value?] that we have to click it and enter it at the same spot on the ball at every time step, and time interval. Every frame, basically.
I [53:05]: M-hmm.
S [53:05]: In order to have the most accurate data, if I clicked on it on one frame on the left of the ball, and the next frame on the right of the ball, we're obviously not going to get accurate motion. Since it's not the same point in every frame, that we have to [click?] the ball. Click on the ball.
I [53:22]: Alright, so with that, um... so with the act of clicking on the center of the ball on each frame, you produce that very slightly curved line up there.
S [53:35]: Yes.
I [53:35]: —in the upper right-hand corner. Uh, which you've been characterizing as... well, what sort of line would you expect if you were in fact—if you were in fact observing uh, net zero constant velocity motion?
S [53:50]: If it was, uh, constant velocity, then the position would just move up in a straight line, with a constant slope, rather than curve upwards or downwards.
I [53:59]: Now is that... that would be true for the plots—the actual plots of position versus time that you would produce like in a computer model, for instance.
S [54:09]: Yeah.
I [54:10]: But can you think of a case in which something that is moving at constantly—at constant velocity, um, might not appear to be moving at constant velocity to a camera?
S [54:25]: Um... if it was moving at zero velocity, I guess. If it was just standing still, but the velocity would be constant.
I [54:34]: Yeah, that would in fact be constant velocity. Um, what I'm trying to probe here is... sources of experimental error. Um, so let's talk a bit about why that—the, uh, trajectory that's calculated up there, why is it—why is it curved? Why is it not a perfectly straight line?
S [54:59]: Probably I think because, you can see a little bit in the video, it does slow down—like, to my eye, it does slow down a little bit at the very end of the frame right there. So it wouldn't—it appears to be moving at nearly constant velocity, but because of the—because of air resistance on the earth, and because of friction forces, it would not be moving at perfectly constant velocity in this case. Because of all the—because of the outside forces of the object. They're always acting upon the ball. But it just depends on what the net forces are doing on the system.
I [55:36]: And you make a distinction between air resistance and friction. So can you explain what you mean by that distinction?
S [55:44]: Alright. Air resistance? Like if you throw a ball through the air, if this had been thrown through the air, it obviously wouldn't have friction, because friction I believe is defined as—it's also defined in terms of the magnitude of the normal force that points upwards—or not necessarily upwards, but... in uh, an average case it may point upwards. Because it's, I think it's defined as the normal force times some sort of coefficient of either static or kinetic friction, whereas air resistance is kind of the force of the air? But there are molecules of the air, they're not really dense, they're pretty spread out. But they're still acting—there's just a lot of tiny, tiny forces acting upon the ball when it—when it would be thrown through the air. So that's different in terms of the force of the floor, and whatever surface it would be acting upon the ball, rather than the forces of the molecules of the air acting upon the ball.
S [56:41]: And... a lot of tiny, like infinitely—and infin—a theoretically infinite amount of tiny forces of these molecules are pushing back on the molecules of the ball.
I [56:49]: Well, and so can you compare—do you have any idea how you might be able to—I'm sorry, there's a bit of audio feedback. Okay, I think we're good. Um... can you think of any way to compare perhaps the magnitude, or, you know, the importance of the force of the air resistance versus the force of contact friction, in this particular case. Do you have an idea of which one might be bigger or, you know, less important?
S [57:19]: I think that uh, the force of friction would be more important in this specific case—because—not one hundred percent sure, oops sorry, I'm not a hundred percent sure what surface this is being conducted—it looks to be a concrete surface, it could be a little bit rougher. There are coefficients of concrete to determine what the actual magnitude of the [inaudible, interacting with message popup on screen]
I [57:45]: I'm sorry, if you don't want the messages to appear in this recording, you can swipe in from the right and turn on do not disturb mode, and that'll suppress all of your notifications.
S [57:53]: Swipe what in from the right?
I [57:53]: Like, two fingers in from the right of the trackpad, and that'll open the sidebar.
S [58:02]: M-hmm.
I [58:02]: Or, or... yeah there you go. And then scroll down a bit, it's kinda hidden up at the top.
S [58:07]: At the top?
I [58:09]: Uh...
S [58:09]: Oh!
I [58:09]: Yeah. There you go.
S [58:09]: Oh, okay! Thanks.
I [58:11]: No worries.
S [58:15]: Okay. Alright.
I [58:16]: Okay, so, lost my train of thought there.
S [58:18]: Okay yeah, and this—to reiterate this—in this case I believe that friction would be more important, because in the normal—in normal air—I guess normal air [laughs], there would be—in this—bleh, sorry. Because it's a small ball, its not, like, we're flying a piece of paper through the air, it's kind of more like a compact object rather than like a spread out one. So, it would, the floor would probably exert more of a force since it's a small object. As well as—it kinda looks a little rough in some place. [There seem to be?] some chips in the concrete, so, the rougher the surface the more friction would obviously, I believe that's how it works, the coefficient of friction would be larger on a rougher surface, rather than the force of a bunch of tiny air molecules.
I [59:16]: So if, so if the ball were rolled over a very rough surface, like sandpaper—
S [59:25]: Yeah.
I [59:25]: —that would have a larger frictional force on the ball, and, um... well, what sort of effect would that have on a ball compared to, for example, if you were rolling it over a very smooth surface like glass?
S [59:35]: Well there's a lot of, with the rough surface, like sandpaper, there's a lot of tiny bumps and ridges built into the material. Like that's just a property of the material. Rather than a very smooth, shiny surface, there's not as many, uh, surface heterogeneity—inherently build within.
I [59:54]: M-hmm.
S [59:56]: So I think—in a very small microscopic sense, if it's a rough surface, the ball might get more easily caught on these little bumps and ridges, and that could slow the ball down, rather than just a smooth surface where [people?] slide on it.
I [60:09]: Alright. And how would that effect the shape of the, uh, position versus time plot in that upper right-hand corner?
S [60:16]: Um, more rough surface would probably slow the ball down more. It would curve... downwards at a faster rate.
I [60:23]: M-hmm.
S [60:24]: And if it was a very smooth surface, it would be more straight. It would be more of a straight line.
I [60:29]: And are there any other potential sources that you can think of—sources of... let me not use that phrasing. Are there any other reasons that might cause that trajectory plot to curve downward.
S [60:44]: There's also, um, I believe the weight of the object could play a part? I think heavier objects, it takes more force to roll at a constant velocity. Like the more—the initial force, not in this time interval. Like in terms of a tennis ball versus a bowling ball. Like you're going to have to push the bowling ball harder to get it at the same constant velocity that you would need to roll—like the same [amount?] to roll it at five meters per second at a constant velocity, you'd need to exert a stronger force on the bowling ball to get it to roll at that velocity. Versus the tennis ball you would need to use less force relatively.
I [61:21]: M-hmm. And how would that relate to the curvature of that, um, position versus time plot?
S [61:28]: If it was a bowling ball, it would more easily curve downward, and have a sort of slowdown. But if it was a much lighter object, it would remain more straight, relatively.
I [61:38]: And that's a separate effect from the effects of, um, friction and air resistance?
S [61:44]: Well, if it—the normal force is tied, I believe normal—not necessarily tied to the weight of the objects. But if this was a constant velocity syste, and the only two forces acting in the y-direction were the force of gravity and the normal force, then they would have—then since the weight is inherent in the force of gravity on the object on earth, the weight of the object would matter in terms of the normal force, and the normal force is also implicit in the magnitude of the friction force. So in a way it is all tied together. But it's also... two separate things. Like you could roll a tennis ball, I'm gonna say on the Earth's surface, and a bowling ball on a smooth surface, and there would be kinda two different effects on it—but it would be the same effects, but there are two different reasons would be slowing down.
I [62:37]: And those two reasons are...?
S [62:40]: The reason is the weight of the objects, and the characterization of the surface, like sandpaper versus [glass?].
I [62:49]: Okay. Um... and... uh, what other sources of error might exist during—in this whole procedure of rolling and taking a video and then, uh, picking out the position points?
S [63:06]: Well I guess one one the sources of error is just picking the wrong time interval. Like you wouldn't want to pick a time interval where the ball is noticeably slowing down, so that's just an error in the person's judgment. But there's also—there could be points where the point mass where you clicked on the ball, it could be a little off? In-between points. Like you could click—like it looks like these two points are more further separated than these two points. That's just another human error, like you could be off by a couple pixels. But that could still affect your data. Hopefully not to an extreme extent but that's also a consideration you have to... delve into, [in your calculations?].
I [63:47]: Okay, very good. Um, thank you so much [student's name], I think that will be all that we need from this interview here. Uh, you could switch back to your camera.
I [63:57]: Uh, so I don't have my uh, I would need to check my schedule and your schedule to see what the final session would be, so I might just get in contact with you about that in the next few days, uh to schedule your final appointment. In the meantime, if you could please take a look at that media release form. You don't have to sign it, you don't have to release it, but if you do—you do have to sign it one way or the other. And then email it to me. And then I will be in contact with you shortly to arrange our next session. And I'd like to thank you very much for taking some time to participate in this research project.
I [64:38]: Alright. Thank you very much, [Student's name]!
S [64:38]: Thank you.

F.7 Student A, Summer 2017 (Post)

I [01]: Alright, okay, so thank you again for agreeing to participate in this interview. It'll go almost identically to the interview that we had a few weeks ago. I hope your intervening time in the course has, uh, has gone well. Um... so if you look at the email that I sent you earlier today, uh, I do need to collect the consent form and the media release form again.
[logistics]
I [03:28]: Excellent. So we can begin. The last time we did this interview, we started out with a—an example program—bleh. An example problem to help get us both acclimated to the think aloud protocol. I don't believe that'll be necessary this time, since we've already been through the first interview. So since we probably have a little more to talk about this time, let's just go ahead and launch right into it.
I [03:53]: If you look through the email, you'll find the lab video that you reviewed at the beginning of the semester. If you still have your notes and your ratings, please don't look at them, because I want you to watch it and rate it again.
S [04:05]: Alright, I don't have them, so...
I [04:08]: That makes my job easier. Um...
S [04:10]: Yeah.
I [04:11]: So please watch it. While you're doing that, share your screen with me, if you would. And, um, follow the think-aloud protocol. When you are—when you're watching

and attending to the video, I know it's difficult to speak very much. But if you're pausing it or thinking about something important, or something important occurs to you, please let me know what you think.
S [04:33]: Okay. Alright, this time, I'm going to um... pull up the documents so you can see what I'm writing.
I [04:39]: That's great, thanks.
S [04:41]: In real-time. Let me... get that... okay. So that's [inaudible] okay.
S [04:57]: So you should—you able to see that?
I [04:59]: Yes, I can see that. Looks great.
S [05:02]: Okay.
S [05:35]: [under breath] Okay
S [06:00]: [muttering] Oh! Must have [stuttered?] after that.
S [06:06]: Intro... okay, it's pretty good, I like how she was all like at the beginning—it is a little busy on this slide, but she did spend a lot of time discussing, but I mean the intro is a very important part of your lab report. She does, um... I like how she does the one, two, three, first I'm gonna record the motion, then I'm going to record the model, then we're going to compare it. So this is a natural split, like, what we do in the lab, that's what we do first, we record the motion. Then that, and that. And it's good—the good—the visuals are good, it's [good to see?] the energy principle's very obvious.
S [06:40]: Does she get more into the fundamental principles? I'm not one hundred percent sure bout that. But I don't think—I feel like that that's not sufficient to introduce like, to completely... introduce the fundamental principle in that way. Cuz I mean, it's good to have the statement, it's good to have the statement right at the beginning, and then repeat it throughout, I mean I'm sure I'll probably find out, like—
S [07:03]: Oh yeah, like I'm pretty sure she, uh... goes back and does more fundamental principles. But... it's the, uh, state your fundamental principle in more than one way—for example, in a sentence like she did here, but also like in an equation, so it would [slide way out?]
S [07:26]: [muttering] signposts... I think they were natural... pretty good... thinking about that. It was—I do give it—her credit for making the fundamental principle very obvious. It was obvious... and she [inaudible] principle... and—I feel like that part over there is not as necessary in the intro since, um, I believe she goes back and does it—the tracker data again.
S [07:53]: In the following... more... discussing a bit and, I also feel like it's a little um... I dunno how—I wouldn't know how to do this better, but I feel like it's a little confusing if you say the model created based on the oscillations of the x position was most accurate, cuz, she doesn't really discuss... that in her third bullet point, so. It's kinda strange how it's, like, if—if I—I completed this lab, so I know what she's talking about, and I know like, what the whole purpose of this is. But if I was just an outside observer, I feel like I'd be well, what do you mean, the oscillations of the x-position? But... I'm sure if she gets into that later and explains that later, and like—but going into it if I was a like, like someone coming into it, like someone who didn't know anything about this lab I'd probably be like "oh, what is the oscillations in the x-position."
S [08:45]: And [all models are?] accurate over time. She didn't do a good job discussing that. I like, I like how she um, laid out results, in the intro...
[assessment discussion]
I [39:10]: Okay, excellent. Let's get right into it. Uh, so this has been uh, the—probably the last video evaluation you're going to do in this year. And you've done quite a few since the last time we had our interview. So I'm going to be discussing a lot about how your attitude toward these videos and towards these ratings has changed over time. And if you could please, uh, switch—
S [39:34]: Do you want me to uh—
I [39:34]: Switch the camera—
S [39:37]: —screenshare, or?...—
I [39:37]: Uh yeah I'd like to see your, uh, face.
S [39:40]: Okay, yeah, okay.
I [39:41]: Okay, excellent. Um, so I noticed that the ratings that you gave uh, were identical to the expert ratings except for one, where you gave a very good and the experts gave a good. Has that been your experience recently in the course, when you were doing the practice videos? In labs three and four?
S [40:01]: Yeah I feel like when I do an error, I either err on the side of [gestures] giving them too much, but there have been times during calibration where I've rated them, like, too poorly. But it's less common that I would rate it more harshly, more common that I would rate it less harsly.
I [40:20]: Well let's talk about that for a bit. Why do you think that is? Why do you think you would tend to give higher ratings, uh, than the experts?
S [40:28]: Hmm. I'm not... well I think something—a document that was really useful to me when I was um, doing my calibration was... a statement there—there was—I dunno. I remember—I don't remember which webassign thing there was, but there was a thing that went through a, "Poor—that's this this and this, Fair does this this and this, Good does this"—I feel that was more, um... personally, like, from distinguishing a good and like a very good, so, I guess like, cuz it would say like, oh it has more than one sort of representation of—I dunno—the principle, is like an excellent, and that distinguishes the excellent from the very good.
S [41:10]: Um, this rubric, it's just like—what is poor, and what is excellent, versus that—I don't have that document one me, like I don't know... I can't access that off the top of my head right now. But it kinda distinguished between a fair and a good, versus a good and a very good, it was kind of more of a spectrum than...
I [41:29]: Was it a document—
S [41:30]: [when I started] using that, then I would get more accurate calibration scores.
I [41:35]: Was it a docu—was it a document provided by the instructors, or was it a student-provide—
S [41:39]: Yes!
I [41:40]: Instructors.
S [41:41]: Yes! I think it was in the... it was in the web—thing. The where you turn in—the submission thing, I believe. It was like a—
I [41:53]: Like, the drop-it or whatever. Post it. Post 'em.
I [41:57]: So in any case, actually, after the interview, uh, if you have a couple of minutes, uh, if you could find that document, that would be—that would be great. Cuz you're the first person to bring that up.
S [42:07]: Yeah, yeah.
I [42:08]: Okay. So—
S [42:09]: I think—it's one that the experts used, but it's not one that they included in the calibration assignment, so I always had to pull it up separately. But it is included from the instructor.
I [42:19]: And with the use of that document, did you find yourself getting, uh, higher calibration grades as the semester went on?
S [42:27]: Yeah. I did feel like that. When I looked at that document—cuz I'd been like, I don't know if it's good or if it's very good, or fair or poor, and I feel like that—it was better at helping me distinguish between like a good and a very good—kinda like the middle scores. So it's better at identifying those.
I [42:48]: Are the middle scores, like, particularly troublesome to make a judgment about?
S [42:48]: Yeah, I feel like with this, it's kind of like, all or nothing. Cuz I feel like sometimes half of these things are advanced, like, the introduction included a problem, but what's like, what is excellent versus what is very good, or like, or what's poor—it's easy to tell what's excellent based on this, but in the middle, I guess it's less obvious, like what makes it poor, or what makes it good. So I guess it—it's just something that has helped me.
I [43:22]: But even with that document, um, you still say that you tend to give higher ratings than experts. You know, when you disagree with the experts, you'll tend to be giving a higher rating. I'd like to talk about that for a bit. Why—why do you think that is?
S [43:37]: I'm not sure. I gues I see more of like, what they've done, and I guess I've seen fewer of these videos, so I haven't had to like... I haven't seen one that's like "wow, it's perfect!" And I guess nothing is perfect, I've never really like gave everything excellent, I don't—I guess... I've personally never seen a perfect video. But...
I [44:03]: I'm sorry you cut out for a second—did you say you never gave excellent... you never give excellents anymore?
S [44:09]: No, like, everything.
I [44:09]: For everything.
S [44:10]: I've never given someone all excellents. Cuz I guess, I haven't seen a video that was entirely excellent. [shrugs] just personally. But I guess... when we see a lot of videos, and have a lot more experience with that, we kinda... have seen stuff that's good, like you've seen students that are good, so you can... see more, like, do more... I dunno, fair? Judgment—like you know more, probably, what is good and what is not good versus someone who's taking the class for the first time, and they're like "they included it, it's good!" but even though someone included something doesn't mean that they did it well, you know.
I [44:45]: And you're talking about, um—so what sort of things are you talking about including in the videos?
S [44:55]: I guess something, like maybe with the content oveall section, like, discussing—discussion contains no physics errors, versus contains major physics errors. If you're taking a class for the first time, you probably won't catch every single physics error, like I didn't catch that one error that the instructor caught. Like I didn't catch that on my um, like I don't think I caught it last time either, so I guess if you're more... familiar with the principles, and if you like worked with this for years and years, and if someone messes up, you'll like catch them on that right away.
S [45:27]: Versus, like you just watch it, and like there could be a tiny error?... that sort of slips over your heard, so you feel like "Oh! It didn't contain any errors that I saw, so I'm going to give it a better score than I would if I could catch this small error, like E versus delta E."
I [45:42]: So, your, uh, you're talking about um, years of experience, and if you have years of experience you can catch these errors. Presumably you're talking about the experts, your physics instructors and professors—
S [45:53]: Yeah, the TAs, the instructors, like... since I assume that the TAs are in graduate school by this point, and have been working with physics for I guess their whole undergrad, and at least part of their graduate. I guess that's what I mean by like years and years. Or like, above that with the instructors, too.
I [46:12]: So—so how would you describe your experience level now, as compared to when you first started this course?
S [46:20]: I think I am more experienced, like I did a cat—I did catch a couple things that, um, I don't remember if I caught them last time, I feel like I didn't. But... I think... you kinda like realize things a little bit more. Like I'm not—I'm not—I'm probably not an expert by this point, but I'm more educated than I was before. About the—certain things. Since I worked through the lab myself, and I have the experience behind be, like I'm probably—when I take Physics 2 in the fall, I'm probably going to have more and more experience to build on.
S [46:57]: And... if you—if you know the concepts well, and if you like study hard, and if you like work at them, then you're more likely—versus someone who's just coasting, and only studied like... rather than, then you'll probably be more able to catch these small sort of things, versus if you—caught—like the core concepts as well. If someone says "Newton's Second Law is every action has an equal and opposite reaction", then some random person who didn't study would be like, "yeah, that's one of Newton's laws, that's fine." But you're like, "but that's not Newton's Second Law, that's Newton's Third law."
S [47:32]: So I guess, I mean, that's kind of like glaringly obvious to me and like anyone else in the course, but... there are little things like that where sometimes you don't notice.
I [47:41]: So you—you're—you're describing, you're comparing, uh, a hypothetical student who is paying a lot of attention and gradually becoming better and better, versus someone who you said was just coasting through. Uh, do you have a sense of how many of your classmates might be paying a lot of attention and putting in a lot of work, versus classmates who are coasting?
S [48:03]: I do think at Georgia Tech a lot more people—like, just the culture, people do put in more work on average. But I guess, during high school that was more of a problem, like poeple would just like not pay attention, and not even care. But at Georgia Tech, I do think more people—there are some people that... didn't have—I dunno—you can kinda tell, cuz they're like on their phone in class, or they're like "I didn't do the homework"... like you can usually tell with that sort of thing, but like sometimes people just really stress, and they just—say they're taking twenty credit hours or something, and they just don't have the time to study, so there's—It's not like they're always slacking, but... usually you can tell if someone's like, outside of class, if they're not really understanding the problems.

S [48:48]: I guess for this class specifically if like, you're in the group hangout and they're like "What's going on, what is this?" like I—"I didn't watch the videos or anything like"—like someone will say "sorry, I didn't watch the videos this week"
I [48:59]: Mm-hmm.
S [49:00]: That student will, I mean... they could always just watch them the next day, but, you can kinda tell who's pulling their weight in that sort of situation. Versus someone who's like on top of things, like "oh yes, this is the fundamental principle we need to use, this is what we're doing..." so you can tell—it's more easy to tell when you're doing your problem work... versus kinda like a big lecture hall, like, I mean you can see people, and you could like hear them and talk to them, but you don't really know what they're doing outside of class, unless you're friends with them.
I [49:30]: Uh, and particularly with the um, peer videos that you've been evaluating, does it come across how much effort the person who made the video is putting into it?
S [49:39]: I—I do think so, it's a lot of—I think... it's like a lot of production... I'm, like, I, uhh. Like I feel like putting a lot of work into producing it, and it kinda shows. Like versus people that just—I have seen some videos that were kinda bad, and it's just people like—I don't know how to describe it, but one was just someone talking into the screen for like five minutes, reading off a piece of paper. I was really confused. I don't know if you like know of that, I don't wanna call 'em out, but I—I guess—it shows versus someone who like has a lot of types of media, like this is what I did here, like here's a screencast, here's like a picture of an equation that I drew, or something like that. So... and also like make sure like everything's polished, and nice—I can tell, I can tell how hard they worked on it, cuz—cuz they usually have good production. I guess that's not always an indicator, you can have like something that looks really good and sounds really good and like can use a lot of different... [inaudible] but can just be completely wrong sometimes. So it's not like the most, but it is—I guess an indicator. Blowing into the microphone, like talking really fast—like did you do this the night before, or something?
I [50:57]: So—
S [50:57]: I guess, you can kinda tell in that regard.
I [51:00]: So, in terms of, judging... effort by production quality, you mentioned uh—it sounded like you were mentioning that hypothetically, high production value doesn't necessarily mean that the content is correct.
S [51:16]: Yeah.
I [51:16]: Have you encountered that? In your experience with these videos, has there been a case where there was high production value but some error, or some omission?
S [51:26]: I think, um... I can't really think of specific examples, but... there have been times where like it looks really good, and they did a lot of—they put effort into it, but there wasn't—there was like, not enough content I feel. Like I can't think of anything specific coming to my mind right now, but I guess you can... like include a lot of stuff, but it doesn't—it's not, it's not always of substance.
I [51:52]: Alright.
S [51:53]: Like you still have to have the model, the principles, the equations, and all that. One thing I kinda have noticed is when people get like, a good score, they usually—like I don't see many people getting like poor in one category and excellent in the other. Like I feel like if you, there's like—it's within like a couple of like—usually like all there scores are within a couple of columns. So I feel like if you work hard in one aspect, you're gonna... it's gonna show in the other aspects. Like I don't wanna, I rarely every see people with scores all over the place. So like, poor, excellent.
I [52:30]: And this is the scores that you're—that you're assigning to people?
S [52:34]: Yes. And from the uh, calibration? It's usually—like not all the time obviously, but usually they're within like three columns of each other.
I [52:45]: Mm-hmm.
S [52:47]: Like, there probably have been like—but also yeah in my own experience, like I don't usually give poor and excellent, like I haven't seen that many people that are like that.
I [52:59]: Okay! Um... so I'd like to take a step back and return to one of our earlier topics. Uh, which was the changes that you've experienced in rating behavior from the beginning of the semester to the end. So you said that one major, um... one major thing that changed the way you approached ratings was your use of this document which I would really like you to find a copy for me if you can, but putting that aside, um... I'd like to talk briefly about any sorts of changes in attitudes or perspectives that you might have had starting the semester versus after you've done a couple of labs.
S [53:39]: Yeah... I guess it's um... I guess it's kinda intimidating at the beginning, like you don't really know what to do, cuz I mean I've written lab reports before, but that's like five paragraphs, here's the intro—this, this, this—so it's like a new form of media, which I guess is—like—I mean I've done all that stuff before, so it wasn't like the actual act of making, it was like "oh man, what do I include, what do I say?" It's useful to write down what I was going to say like before I actually said it, which I get—it sounds really obvious now that I say it, but the first like I think two labs, I just kinda spoke about what I said, I didn't really write anything down. But for the last couple, I like wrote down what I was going to say and then read it off of a piece of paper. I guess a voiceover. And I feel like that really helped, like, in terms of knowing what to say.
I [54:30]: And what about—
S [54:31]: But I guess—
I [54:31]: Um. And what about in, uh, reviewing other people's videos?
S [54:37]: Yeah, I think you kinda—when you look at stuff, writing like that, it personally helps me understand labs as a whole better. So if I believe I completed the lab with more effort, and more like time into it and put more thought into it, it helps me kind of understand other people's better. To see where they kinda—not where they went wrong, but also like, what they did and like what went right. And if they like completely missed one thing, and I'm gonna—obviously I'm going to notice, I'm like are you thorough, I would like be very thorough in doing the lab, hypothetically.
S [55:13]: So I guess once you like compl—I think once after you complete the lab, it's a lot easier to kinda speak about the lab.
I [55:20]: Mm-hmm.
S [55:21]: Since you've done it yourself. But... yeah.
I [55:24]: And the more labs you complete, does this uh, does this confidence you get from completing a lab carry on to later labs?
S [55:33]: I do think it's kind of like, like... when you get—I guess, this is like [inaudible] motivation. But... when you like get the grades back from a lab, and if you do bad, it's like—spurs you do to better like next time. Or it can be discouraging sometimes, like oh man what did I do wrong, I thought I did well... situation. But if you get good grades on a lab, like I see an A on the lab, hypothetically, I'll know what to do for the next time. Like, what did I do well—I know what to do better.
S [56:05]: But that's one of the ways, but also... I guess yeah, once you start completing labs, it gets easier over time to keep completing, like the labs. Cuz... and—[inaudible] you know—like you keep watching other people's videos. I feel like it is helpful in knowing like... like seeing where they went wrong, and what they did right. Like I've definitely seen people's videos [inaudible] keep that in mind next time. So I think that is useful in terms of that.
I [56:34]: Okay. Um... and... do you think that, uh, so judging by the feedback that you've gotten in your videos, uh, do you think that your peers who are rating your videos are undergoing a similar sort of evolution? Do you feel like the ratings that your getting are fair, is the feedback useful?
S [56:55]: Yeah. I do think the feedback's useful. I mean. A couple labs... it was—it was kind of—it was not like—superficial, but more—less deep, and more like "oh, well, you missed like"—I didn't misspell a word, but like—but oh your—someone told me my handwriting was messy. I was like, [wounded] "ohhh." [laughs] I was like honestly, my handwriting? The next time, I was like, they could have rated me like poor because my handwriting was messy. So I feel like, like not—sup—more superficial like that, but I do feel like it went... over time, they were like, oh well you could've explained your fundamental principle differently, or like you could've done... [inaudible]. So it did kind of go from—I wouldn't say superficial, but it was more focused on like... physics aspects. But like.
I [58:00]: So it was more focused on physics aspects when? I'm sorry.
S [58:00]: At the end. At the end. Like in the last—the feedback for the past few labs, the last ones.
I [58:07]: Do you think that's a good thing? What does that mean [to you]
S [58:13]: [I do I do]. Like you should be focusing on the content of the lab, that's probably one of the most important parts.
I [58:21]: And—is that also a similar... have you seen that trend in yourself as well, the focusing more on physics content at the end of the semester?
S [58:30]: Yeah, I feel like what I've been focusing on more is like oh, well, how do I make it very obvious to the viewer like what I'm trying to like get across and the connection, trying to make the connections between the models and the principles very obvious. So I do think that content is something that matters a lot.
I [58:49]: And when you were evaluating the video at the beginning of the hour, at several points you would say something to the effect of... "new people who are just seeing this for the first time, they wouldn't know what she's talking about". Um, can you talk to me a little bit about what the intended audience for these videos is, and how you think about making this videos for an audience?
S [59:12]: I guess—I guess um... I make them for someone—I guess personally, I would make them for someone who like—is in the class and is grading my video. Like [inaudible] a lab report I guess—your audience is your TA, or the other students. It's similar, but I also want to make it obvious like, if you know these concepts, like, you can figure this out. Like... and the random person walking in off the street, I don't know if they'd be able to understand my video, but also... explain the most simple of things like, "Oh, this is my Tracker software." This—I would like run out of time in my video just explaining the simplest of things. In terms of like—cuz I want to get more into like, the content, the what if questions, the what does it mean questions. And I wanna discuss more about the physical aspects of it, rather than like what all of this is, Newton's Sec—I mean yeah that also is very important. But I guess, for like if I show this to like, I dunno, one of my family members who hadn't been taking this class, they'd understand most of it but they'd probably be a little confused.
I [01:00:29]: Mm-hmm.
S [01:00.14]: For part of it, I guess. Cuz I don't like, I guess, there's just not enough time in the day to explain like whay Python is to your viewer.
I [01:00:38]: Right.
S [01:00.28]: Every single time. Yeah.
I [01:00:41]: Thank you so much, so I'm going to call that the end of the structured section of the interview. There's one last thing—we started at about eleven after, so we have a few more minutes. If you look in the email I sent you, there's a ball rolling example video. Uh, and what I would like you to do is just watch that video through to the end, it shows some video data of a ball rolling across a surface, and then Tracker data from that same video. And briefly, I would like you to explain to me the Tracker findings. Explain why the trajectory in Tracker has the shape that it does.
S [01:01.29]: Okay, do you want me to screen share, or [just watch]
I [01:01:19]: Yes, please.
S [01:01.05]: [under breath] okay, pushed the ball... it's a tennis ball... from left to right. Across the screen. [I think?] there's a lotta—okay
S [01:01.10]: So then they start—they start it... it's after her hand leaves the ball. And it kinda... x-position... is, it's not moving in the y direction at all, I believe. Or at least, not significantly. Kind of... rolling, it's getting slower. [Go to the?] video, and it stops right before the... um, this part, the edge of the ball leaves the frame. So it starts after her hand leaves the first—the person's hand leaves the ball. Forces acting upon it other than the force that slows the ball down cuz, eventually the velocity decreases. See, it kinda curves a little bit. Not to an extreme extent, but... I believe it slows down cuz of probably air resistance, and friction... forces. It does look like a concrete surface, there are some chips in the paint, so I guess—it looks to be a pretty rough surface.
S [01:02.13]: And a tennis ball's kinda fuzzy, so it's also kinda rough. So I guess the friction forces are probably the thing that slows it down the most.
I [01:03:07]: And so if the roughness of the surface is a signficant factor, how would you expect this data to change if for example you rolled the tennis ball over, uh, glass, or over an ice rink?
S [01:03.12]: Not slow down as—like it wouldn't curve as much, like that's not an extreme curve, but it would be more of a straight line, and it would move at more of a constant velocity in this timeframe. If it was glass, cuz the force of friction—the coefficient of friction would be... less. Wait—it—it would have less friction.
I [01:03:42]: And, uh, what if instead, um... what is the role of friction in this data, versus some other situation where perhaps we had a cube that was slid across the floor, instead of having a ball rolling across the floor? Can you talk about the role of friction in those two different cases?
S [01:04.05]: Uh, like a cue stick?
I [01:04:09]: Uh, like a hockey puck. Say that we had a—

S [01:04.24]: Oh.
I [01:04:12]: Yeah. Say that we slid a hockey puck across this surface, versus a ball rolling across the surface. What is the role of friction in both of those two cases?
S [01:04.12]: There's less of a surface area touching the ground on the ball, versus the puck is more flat.
I [01:04:30]: Mm-hmm.
S [01:04.12]: The surface area that is touching the ground... so I believe it would probably be more prone to slowing down by friction if it was flatter. The ball is like a... theoretically it's a spherical object. So it would... since there's less contact between the object and the ground in the ball situation, there'd be less of a force of friction acting on it. If, if it was the same surface it was sliding across, at the same speed.
S [01:04.03]: And the puck would, since it has more—it's more in contact with the ground, there would be more chance for the ground to exert a friction force on the puck.
I [01:05:05]: Right, so the, um... if I'm summarizing you fairly, I'd say that the ball experiences less frictional force because the point of—the contact area is less than for a puck, which has a much larger contact area.
S [01:05.07]: Yes.
I [01:05:19]: Um, is there anything else that might cause a difference in the role of friction between the rolling ball and the sliding puck?
S [01:05.01]: Um... it could be the texture of the ball itself. I know tennis balls are kinda fuzzy and rough. And hockey pucks are more smooth?... so I guess that also—the material of the object itself could also, as well as the material of the floor itself, could affect the... [physics of the?] system.
I [01:05:47]: Okay. Thank you so much! I think, I think that's gonna do it for us for this evening. Um, the last form that I emailed you a little bit earlier is the subject payment form. Now, are you coming back to Georgia Tech in the fall?
S [01:06.08]: Yes.
I [01:06:05]: Are you going to be an employee of Georgia Tech? Like on a work study program or something.
[logistics]

F.8 Student B, Summer 2017 (Pre)

I [03]: Alright, and we're going. So, thank you for agreeing to participate in this interview. Uh, before we begin, I emailed you a couple of forms, uhh, just a little while ago, uh, before we can actually begin the study itself, I'll need you to sign the consent form.
[consent is acquired, technical difficulties resolved]
I [06:38]: Let's begin. So first, I'll tell you about what we're going to be—uh, doing during this study. The goal of this study, uh, is to try to understand what sort of changes uh, in physics concepts and behavioural competencies that our course effects on students, which is why we're scheduling for one session at the beginning of the term, and another session at the end of the term, just to see how your opinions and attitudes changed over the course of the semester. Um... uh, this study is gonna be basically identical to the study at the end of the term.
I [07:12]: Uh, we'll begin by, uh, a think-aloud protocol, um, and... we'll begin by walking you through a think-aloud problem for practice. So the way a think-aloud protocol works is that I'll ask you to examine something, or solve a physics problem, and you'll do it as you normally would while narrating to me the thoughts that you're having in your head as you do it. Your thoughts don't need to be coherent, they don't even need to be in complete sentences, I just kind of want to get a little peek inside of what's going on as you're solving the problem. Um, I won't ask you any questions during the think-aloud protocol, but if you're a little too quiet for too long, I might prompt you to keep talking. And that'll be about all the interaction I do with you during that time.
I [07:59]: After that, we'll go through a—one of the videos I emailed you was a lab video, much like the lab videos that you evaluated in this past week, and I'd like you to evaluate that lab video, along with the rubric, as long as you were doing your lab evaluation excercise.
I [08:18]: Uh, after you're done with that, I'll show you what the expert evaluations were, and I'll ask you a few questions about the video, and about your evaluation process.
I [08:27]: Finally, I'll be—there's a very short of a ball rolling across the screen, uh, that I emailed you. I'd like you to examine that, just provide an explanation to me of what the physics is—I'm sure you're used to that sort of procedure by now. And then we'll have some followup questions and a final interview. We expect the whole session to take about an hour. Um... so with—if that sounds alright, let's go ahead and begin with the, uh, warm-up question.
S [08:59]: Okay.
I [09:00]: Okay. Um... so, again, a think-aloud protocol, so you can write things down, you can use a calculator, it's just if you produce any documents, if you could send me a scan or a picture afterward, um...
S [09:15]: Okay.
I [09:17]: This is a two-part estimation question. So, what is the volume of the room that you're in right now, and if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for the room to fill with water?
S [09:34]: Okay. So you want me to, like, this is where I say what I'm doing out loud?
I [09:41]: M-hmm.
S [09:41]: So, the volume of this room I'd say... the height is probably twelve feet...? The length... is probably... probably ten feet. The width... probably twenty. Yeah. Or... Probably like eighteen. [inaudible]
S [10:24]: Just multiply all that... mmm... that'd be two hundred—two thousand one hundred and sixty cubic feet. A kitchen faucet is... if you turn it on, like all the way, probably could fill up a gallon in... a minute. Mmm... I think there's... I can't remember from my engineering classes, but I think there's like twenty-seven gallons in a cubic foot?
S [11:09]: Does that sound right? I dunno... so if that's right, then... two thousand one hundred and sixty divided by twenty seven... wait, no, there's twenty seven foot cubed in a gallon. Wait! No, that would be... [laughs]—so that would be eighty gallons, so it would take like an hour... and twenty minutes. That seems kinda low, but... maybe? If your kitchen faucet is... quick. Yeah. Hour twenty minutes.
I [11:47]: Okay, thank you very much. So that's about... that's about the level of narration that we would like, so if you pause for a moment to just think—if you're doing something visually in your head that's difficult to express in words, but keep up the narration, I think you're doing a good job with that. That's how the think-aloud protocol works, um. The next part of this study is going to be a video evaluation, and so, uh, please open up that lab video that I emailed you earlier, and if you could also please switch your Hangouts session to present your screen rather than your camera. I'd like to see what's going on on your screen while you're watching this.
I [12:30]: And before you do that, uh, make sure there's no visible sensitive information on your desktop, because that's going to get recorded.
S [12:40]: Okay, let's see... mmm... I'll just delete like my emails...
I [12:46]: Right.
S [12:48]: Is it the one with like the peanut butter looking thing, or like the tennis ball?
I [12:52]: [under breath] the one with the peanut butter looking thing?... oh, yes, it's the one with the peanut butter looking thing.
S [12:58]: Okay. Yeah. Alright. Mmm... share screen's not on this for some reason. It's just a little chat bubble.
I [13:08]: Really?
I [13:09]: [talking over each other] yeah it's the—
S [13:10]: Okay yeah—
I [13:10]: it's the veritcal ellipses, yeah.
S [13:15]: I can on—oh, I can share the whole desktop thing?...
I [13:18]: Yes please, because as you're doing this video, as you're evaluating this video, please keep the rubric up and please type out your ratings and comments. I would like to collect those at the end of the study.
S [13:31]: Okay. Um, did you send me the rubric—oh it was like the actual attachment, let me [unintelligible] at that
I [13:41]: Unfortunately it's not editable, it's just a screenshot of the rubric, so you'll need to type out, you know, "Item one, rating, comment, item two rating comment" in a text file or something.
S [13:55]: Okay.
S [14:17]: It's like really small.
I [14:17]: Is it. The video, or the rubric?
S [14:19]: The rubric.
I [14:20]: I thought I sent it at full size... uh one moment, I apologize...
S [14:25]: That's okay, it's just I can't zoom in because it gets blurry.
I [14:29]: Right, I think I know what's going on.
I [14:45]: Okay, I've sent you a full size version, uh has it arrived?
S [14:53]: No...
S [15:36]: So you just want me to watch the video?
I [15:36]: Uh, yeah, so watch and evaluate the video as if you were doing it for a lab exercise.
S [15:43]: Okay.
I [15:47]: And if you—yeah excellent, thanks for sharing the screen.
I [15:55]: Wonderful.
I [15:58]: And now, when you—when the video is actually playing, and you're listening to it, I understand it's fairly difficult to speak your thoughts aloud, uh but while you're making your judgments and evaluating, please do continue the think aloud protocol.
S [16:13]: Okay.
S [16:35]: I think his introduction is kind of short. But he like, said what he [inaudible].
S [16:55]: Yeah.
S [18:01]: He did a good job in explaining the initial conditions.
S [18:07]: The audio's like, really clear.
S [19:08]: [Laughs] okay that really didn't make sense.
S [19:45]: It's not the initial position every time because it's a loop, but whatever.
S [21:18]: That's very [impatient?]. Okay, that's it.
S [21:21]: So, I'd say that his introduction—I don't know. It didn't really like, I didn't really like it, it didn't really state the results, um or like preview the whole lab, it was just like one sentence. But like the audio was really clear and I was like never confused like on the structure itself. Um, I'm pretty sure he's—he did saythat it was like the Newton's second law, whatever, and he discussed like the main physics ideas but it was kind of weird when he said like—the forces are invisible, like or hard to determine, like they're NOT—I dunno, that didn't make sense, cuz like at the end he said it was friction, so I don't know why he said that. Um... and I think he spend like two minutes like explaining the python code, because I also feel like he didn't really know exactly what he was talking about, like he did good enough I guess, but I dunno, that was kind of sketchy. Um. But like I'm pretty sure his code was right and his Tracker was good. Probably didn't need to show the whole Tracker video at the speed it goes at, but whatever. [inaudible]
S [22:32]: Um... he, I know he answered the what if, like the axis was flipped—I forgot what the what does it mean question... I forgot what that was for this lab. Um... the phy—no physics errors, I think... I dunno, when he said like the forces are invisible and hard to determine that was kind of a weird error to say, I dunno. That was just really weird to me. And the production delivery, I would say that was... pretty good. Ver—at least very good.
I [23:01]: Okay, so could you please uh, put specific ratings. Yeah.
S [23:07]: Oh! But... oh. Is it okay if I did that—I thought you were going to tell me to—would you like me to re-share it?
I [23:16]: Yes please. And if you could please write up or type down actual specific ratings between for—between poor to excellent for each item.
S [23:24]: Okay. Can I circle them on here.
I [23:27]: Uh yeah, you can circle them, you can type them in a text document. Whatever. Whatever's easiest.
S [23:45]: Okay. Um. I would say that this one... I'm going to give this one a fair because he didn't really do a good job in the introduction.
S [24:04]: And this one... I'll say this one was very good. He explained it... it was kinda sketchy and in the middle but I think [inaudible] good job.
S [24:23]: Mmmm...
S [24:40]: I'd say this one's very good too.
S [24:55]: I don't know if he answered the what does it mean so this is good. If he did answer the what if, even though he said the forces were invisible [inaudible]. And then [inaudible]
S [25:12]: I'll say this one was excellent. The video's been... Is that good?
I [25:20]: That's good.
S [25:22]: Or do I need to...

I [25:22]: So I think what you've, um, what you've spoken as the comments for this case. You gave a very clear explanation for why you were giving each rating, which is what I'm interested in.
S [25:34]: Okay.
I [25:36]: Also, if you want to turn on do not distrub mode, that'll keep those notifications from popping into the recording.
S [25:44]: Like on my phone?
I [25:45]: Like on your computer—if you click on that right icon, the three horizontal lines, the very top right of your screen.
S [25:52]: Yeah?
I [25:53]: And then pull that down a little bit, you can turn on do not disturb. Yeah way up at the top—scroll up to the top, there.
S [26:01]: Oh I see.
I [26:02]: You can turn that back on—
S [26:05]: I was wondering how to do that!
I [26:08]: Yeah it's not the most discoverable thing, they literally hide the button. But, in any case...
I [26:11]: Okay, so excellent, those are your ratings, so now like you would be doing with your laboratory practice exercises, these are the instructor ratings for that same video. I just emailed them to you. And if you could please take a look at those, and go over them, and tell me what you think about it.
S [27:08]: Alright we'll see [inaudible]
S [27:17]: So he said poor, and I said very good, mmm... fair... okay, well, we were only like really off on content models. I thought he did say that it was the newton's second law, in the very beginning. I guess it wasn't a great explanation, but I wouldn't say there was no explanation but whatever. Yeah, maybe.
I [27:50]: Okay. Well, so let's—now I'd like to have a little discussion about that video exercise and your ratings and the expert ratings. Uh, you can switch over to video of your face now.
I [28:05]: Okay, so, um very broadly, uh how would you say that this video compares to other videos that you've watched, both for the practice and calibration and for the peer evaluation part.
S [28:18]: I'd say that it was pretty normal. Like... there were some that were better, but it was... it was as good as some of them, that's all.
I [28:29]: I—
S [28:30]: So [interrupting]—
S [28:34]: Yeah. I'd say it was good as a lot of 'em I saw.
I [28:35]: Have you had a generally favorable or disfavorable impression of the peer videos that you've been watching? Could you even say that you have a general [talking over each other]—
S [28:45]: Favorable.
I [28:45]: Oh okay a favorable impression. Could you tell me a little more about that?
S [28:47]: I don't know. I just... uh... I'm like a really easy grader, so like I always, I hate giving like below good, it just makes me feel bad. And like—
I [28:59]: Tell me a little—
S [29:00]: And—
I [29:00]: Let's talk about that, why does it make you feel bad?
S [29:02]: I don't know. Cuz I know—like unless I can blatantly tell someone just did not care at all, I just feel like you should get good for at least trying. Unless I can tell like literally they were just like "let me do this in like quickly five minutes" and I can tell that, then I'd say I gave like one person a poor, on like a couple things when I was grading last time.
S [29:26]: But other than that, like—and like also since I know a lot—or I don't know a lot of physics, but since I made a video too, and I can pretty much know what they're talking about, even if they don't like explicitly say it, like sometimes they just imply it, I guess you could say...
I [29:43]: And so how does that effect the way that you... create these videos? You know... who's the audience that you're creating videos for, and how do you decide what to leave in, and what to leave out?
S [29:59]: Um, just like when I was making my last video for example, like I made sure to like really explain how I got my b, because I knew like a lot of people didn't calculate their b the same way I did, because I used a polynomial regression line in Excel, so I made sure that I really explained things that might be different and just, like, if i know... that... um, like it's something that like pretty much everyone has, I make sure not to spend as much time on it. I just don't put it in my video, obviously.
S [30:31]: That's why I don't really understand the point, like, of the introduction, because like we know what the problem is, and we all probably got like around the same outcome, but... yeah.
I [30:42]: Uh, and very quickly what you—can you explain what you mean by "the b" you measure?
S [30:49]: Oh, the b, like b proportionality constant for um... the drag.
I [30:57]: Oh, for the drag—so you're talking about your latest video, the lab 2 video.
S [31:02]: Yeah that's what I'm talking about.
I [31:04]: Okay, excellent. Um...
I [31:05]: So tell me why do you think uh, the... instructors put the introduction on the rubric? What were the instructors expecting to see in putting such an emphasis on the introduction?
S [31:19]: Just, I guess... I could like like it could really make it clear that you know what you're talking about in the lab, and just like quickly sum it up, so it'd be like easier to understand, I guess.
I [31:33]: Well, so—very generally, what do you think, um—so you saw the instructor evaluations, uh, compared to your evaluations. Can you tell me a little bit about how your ratings compare to instructor ratings in your experience?
S [31:49]: Um, I, oh you mean like—not just this video, all of 'em?
I [31:54]: Yeah all of the ones that you watched for practice, yeah.
S [31:57]: Um... I'm pretty sure they were like, the same, and like mine was just like one ranked higher, or something. I don't think—I don't think mine were ever lower, I don't think.
I [32:11]: And why do you think that is?
S [32:13]: Um [laughs] like I said I feel like if I can tell they tried, I'm just easy—I'm easier on them.
I [32:20]: So would you characterize the instructors as being hard on students?
S [32:24]: Um, no, but like—and also they know like exactly what they're looking for, and I'm just like looking for what I'm looking for in my brain or going off of what I did, when maybe what I did wasn't right or somet—or not not right, but not as in depth as the instructor wants. I meant the instructors at Georgia Tech are like harder than at other colleges, but I just think that's why my ratings are like... higher than the instructors. Cuz also I don't know exactly what they're looking for, only the instructors do. Even though I had the rubric, but still.
I [33:01]: And do you... how do you expect that might change, you know as we get closer to the end of the summer, and you have more experience grading these videos?
S [33:10]: Um... I think... mine might get like a little but harsher, I don't know, because like I did the calibration whatever, and then I started doing with the peer videos, and once I got on the third peer video, I realized—okay, this guy actually didn't have an introduction, I went back and changed one. I could see me doing that more maybe in the future.
I [33:36]: So and—was it... would you describe that as an attitudinal change, like you had a more critical—or you know, you had higher standards, or is it more like a perceptive change, that you're able to notice more—more flaws.
S [33:50]: I'm able to notice. I'm just not a very critical person.
I [33:53]: Mm-hmmm. Critical in the sense of having negative opinions, or critical in the sense of paying attention to things?
S [34:04]: Having negative—like I'm always trying to find, like, how can I give a better grade, than like worse. Like I'll be like, "oh, he didn't mention this when we was supposed to, but he kind of mentioned it at the end, so it's fine." You know?
I [34:19]: Do you have much, uh, experience doing peer grading in other courses? Have you ever done peer evaluation like this before?
S [34:27]: Yeah, like, lit. Literature and stuff. But not science-wise.
[off-camera disturbance]
S [34:41]: [laughs] okay.
I [34:41]: So... but—so tell me little bit, actually, about your peer evaluation experience during literature. Do you find yourself having the same attitudes, like critical, uncritical...?
S [34:52]: Uh, I would say I'm definitely more critical than I am in science, because I know a lot about literature and I'm really good at it. But like, still not in a mean way, like—or—I'm not saying like when... I don't know. I just feel like I feel so bad when I give like a fair or a poor... like I just, I don't know. Like I have to really have a concrete reason on why I did that. If I don't feel like that, then I just put a good, or maybe fair if... I don't know.
I [35:29]: [yelling in background] Do you need to take care of that?
S [35:31]: No, it's my mom yelling at my brother.
I [35:34]: Okay, okay, cool.
S [35:36]: [laughs]
I [35:37]: So do you—so you think that, uh—so it seems that your—mh. It seems that you're making a connection between your content knowledge, you know how much expertise you have in physics or in literature, and the—and your attitude toward—not your attitude, but the ratings that you give. That if you have more knowledge, you might be able to give lower ratings. I don't want to put words in your mouth but that kind of sounds...
S [36:04]: Yeah. It makes me feel like pretentious if I give a poor, because who am I to say if it's poor or not, you know?
I [36:13]: Ah, so you don't have the um—you don't have the authority to give poor grades to people?
S [36:16]: Yeah. Yeah. I guess when it comes to it, it feels like that.
I [36:21]: Very good. Um... so what were your... so in that general context, um, what was your expectation, your general expectations of how this peer evaluation, uh, process would be like. When you first learned that you had to do it in this course.
S [36:40]: I thought that it would be like it is. Like I figured it would be similar to that, like watching the videos, and having a rubric, and doing that. Maybe—not like having to add comments, but... you know. It wasn't like surprising or anything.
I [36:56]: M-hmm. And would—how would you describe your general level of, um, confidence in the accuracy of grades that come out of this system?
S [37:09]: Mm, I'd say like... at like... if everyone's grades are like near each other, then fine, but if there's like one fifty, one ninety, one sixty, like then it might be a little like, "what is going on?". But I would say... I would say they're pretty fine.
I [37:32]: Do you expect that your peers have the same kind of general approach to peer grading that you do?
S [37:41]: Um... I dunno. Maybe. Because like, I dunno. Some people—some students at Georgia Tech are pretty confident in themselves. I'm not saying I'm not confident in myself, but... I'm trying to say like, they... some students, I dunno. Maybe.
I [38:03]: Um... alright, I think that's about all that I have for now for this part of the interview [cat jumps into view] Um, please forgive me.
S [38:16]: [laughs] That's a really great cat, it's like really fluffy.
I [38:21]: Yeah she's an incredibly fluffy creature. Um, so for the final part of the interview, I'd like you to open that last video with the tennis ball. Um... give it a watch, uh, tell me what you're thinking—follow the think-aloud protocol to begin with, tell me what you're thinking about it, provide an explanation as if you were explaining it for a lab 1 video, uh, and then we'll have a discussion about that for the last, uh...
S [38:46]: Wait, what do you mean explaining it?
I [38:48]: Like, tell me—describe the physics going on. So it'll be video footage of a ball rolling, and then video footage of the tracker analysis that produces the position versus time plot. And I want you to explain to me the physics as if you were preparing an explanation for the audience of your lab videos.
S [39:09]: Okay, so do I watch it first and then say it?
I [39:14]: Watch it and think about it and then just kind of generally—I'm not going to tell you what order you need to do things. But take—I'm giving you the direction to watch the video and explain to me what's going on.
S [39:25]: Okay.
S [39:29]: She pushes the ball... it looks like it goes in a pretty straight line. I'm guessing this would be the momentum principle. Okay. Now she's tracking it... going in—
I [39:46]: And I'm sorry, could you switch over to sharing your screen?
S [39:51]: Sure.
I [39:51]: Thanks.

S [39:59]: Okay. Still dragging it... kind of in a curved line, so... that means that—and um, and this is from lab one, like computational model would have shown the straight line, the tracker shows a little bit of a less straight line, so... this would be the momentum principle. Except not with a Fnet of zero, um, and then... in reality, the Fnet wouldn't be zero, and would be the full momentum principle. So like, that's like the physics, this would be used, and then tracker shows that there is... not an Fnet of zero.
I [40:43]: Okay so, tell me, tell me—let's talk a little bit about that Fnet not being equal to zero. So what—
S [40:50]: Yeah.
I [40:52]: Tell me your line of reasoning for coming to the conclusion that there is no—um, that there is in fact NOT Fnet equals zero.
S [41:00]: Because if Fnet did equal zero, this would be completely constant. Because you have to have an Fnet equals zero to have like, constant, speed. Which it does not have that. So.
I [41:14]: And so what force or forces might be contributing to the nonzero net force?
S [41:22]: Um, friction of the—looks like that's like a garage floor or something, so—or maybe... maybe like a patio floor. Could be concrete, which has all those little grains in it. So that could definitely cause a lot of friction, especially with a tennis ball.
I [41:39]: And so we—sorry, continue.
S [41:43]: Cuz it's like the tennis ball has hairs, or you know it's like fuzzy, and concrete has like sticky little dots in it, so that would cause a lot of friction, and that would make that not so straight.
I [41:56]: And so if the tennis ball were rolling along a different surface, like say glass or something, a smoother surface—that curvature wouldn't be so prominent?
S [42:07]: Um, yeah. Like if it—if the... surface had less friction, then I think it would be less curved.
I [42:17]: Uh, and are there any other forces that might contribute to that curve? Besides the contact friction?
S [42:23]: Mmm... [yelling in background] I guess... I mean it's going in the x direction, so I don't think gravity would have that much... um... no. I don't think so. I don't know. If I'm—I'm sure there might be, but I don't know, I can't think of them.
I [42:47]: Alright, real quick, could you just run through a list of what you think the, of what the individual forces acting on this tennis ball might be.
S [42:58]: Um... well... I think just... friction. Because—gravity, I mean gravity... could be, but for only looking at the x-direction, then I don't think that gravity would calculated, like in the Fnet of the ball. Cuz gravity's in the y-direction, usually.
I [43:24]: Okay. So, um, if you were drawing a free-body diagram of the tennis ball, uh, there wouldn't be a gravity force on that diagram?
S [43:36]: No... actually...
I [43:40]: And if you want to draw such a diagram, please feel free to go and do that.
S [43:45]: No, yeah, I guess gravity would be, cuz like if there was no gravity then it would be floating. So yeah, gravity is a thing. And the friction would be like pointing to the left, and then... like whatever pushed the ball... wouldn't have an arrow but like that's how it would be moving.
I [44:08]: Hm. Now why wouldn't the push have an arrow?
S [44:11]: Cuz it's like not a force—it's not a force on it anymore. It's just a one-time thing, then... it doesn't happen anymore, I guess. You don't include it in the body diagram.
I [44:22]: Alright. Um... and so... getting back to the uh, the curvature of the position vs time plot, uh, is the only possible explanation that there are, uh, that there's a nonzero net force on the ball, or might there be other ways to get a curvature in that position plot?
S [44:43]: Um... maybe if like, the tracker—or like it moved not in a completely straight line, and the tracker picked up on that. I think that could make it, cuz it's—cuz it would like move it, a little bit because it didn't move in a completely straight line.
I [45:05]: Okay. Um... and, what about in the process—uh, tell me a little bit about the process of actually picking out the points, uh, of where the ball is.
S [45:20]: Uh, I would just alwyas like use the center of the tennis ball, if it was me doing it.
I [45:26]: And how do you tell where the center is? I know that sounds like a dumb question, but, everyone has a different technique.
S [45:30]: Um... either I like use the autotracker, and then like adjust it if it's like not on the center, but it usually gets it like perfect.
I [45:40]: M-hmm.
S [45:40]: Or I just try to make my best guess of the center.
I [45:46]: Alright, very cool. Um, so those are all of the follow-up questions that I have. Uh, I'd like to thank you for your time. Uh, interview ended about the right length, a little bit shorter than the hour we had scheduled, but we'll still pay you for the whole hour. Before, uh, the—before we do the payment though, we do need to do the second session which will come at the end of the semester. And once I get my calendar up, I'll contact you to schedule that second session. Um...
S [46:23]: Okay.
I [46:24]: It'll be an hour at the end—toward the end of the semester, after you do your lab 4 video. Um.
S [46:31]: Okay.
I [46:31]: So keep an eye out for that, I'll get in contact with you over email. And again, I'd like to thank you for your time, I think this was a very productive interview session.
S [46:41]: Alright, thank you!
I [46:42]: Thank you very much, have a good day.
S [46:44]: You, too.
I [46:45]: Bye.

F.9 Student B, Summer 2017 (Post)

I [03:29]: So let's begin. So thanks again for agreeing to participate in this interview. I hope you've had a um, a good experience with the rest of the course over the whole summer. The purpose of this interview here, it's going to be very similar to the interview we did at the beginning of the semester so that we can judge what sort of changes you've made, you know, since the last time we've talked. It'll start out directly with the video evaluation, it'll in fact be the same video you evaluated the first time, so if you still have your evaluation, if you still have the rubric you filled out, please don't look at that, I'd like you to look at it with a fresh pair of eyes.
I [04:11]: Um, during the video evaluation we will be using the think alout protocol, but I don't think it'll be necessary to do a practice problem for that again. We'll go straight into the video evaluation. Then after that, we'll have a bit of a discussion, then after that I'll ask you to review a physics observation video, uh, summarize the physics for me, and then we'll have a discussion about that. All told, we plan for this to take one hour.
I [04:38]: Um... so with that, if you could please share your screen with me, uh, and view and evaluate the lab video that I attached to that last email.
S [04:58]: Is it the ball rolling one?
I [05:00]: No, it's the other one, it has a random string of numbers and characters as its filename.
S [05:08]: Okay
I [05:49]: And while you're watching the video, I know it's difficult to narrate your thoughts out loud, you know, when you're trying to listen to something, but when you're filling out your evaluation, uh, when you're pausing the video and... clicking back and forth on it, uh, those are the times when I would like you to be speaking aloud.
S [06:04]: Okay.
I [06:14]: Let me know if that rubric is too small for you. Yeah, it looks like it is.
S [06:17]: Yeah it's pretty small [laughs]
I [06:19]: Let me email you a larger version right now.
I [06:32]: [under breath] I don't know... [aloud] I don't know why it does that...
S [06:36]: Yeah it did that last time, too.
I [06:37]: It always—cuz I select Image, Actual Size... in any case, this should work, I just sent it again.
S [06:46]: Okay.
S [06:57]: Yeah that's much better.
I [06:57]: Okay.
S [07:12]: Okay I'm going to start.
I [07:12]: Wonderful.
S [07:20]: Okay there's pretty much not an intro at all. They just like said... their purpose, which really didn't say anything, didn't state the problem, didn't make a statement of result, or like preview the major sections or like have any, like... I dont—I really like—I know it's not like necessary but I like it when there's a powerpoint or something so you can see, like, oh this is introduction or whatever and he didn't have that so I'd definitely say, like, poor for that part—organization structure.
S [07:49]: Like, just on like in—it might change, but we'll see.
S [08:13]: Okay, so "our axes are straight"? Is... that's... I kinda wouldn't have said it like that. And... um, he just said like, "our model is the peanut butter jar", I think that's what he said, that's not like a model. They mean like, what's the like, principle or like model—like the physics model you're going to use. Um...
S [08:39]: And he just said like the system's a peanut butter jar. Ah, I dunno.
S [08:56]: Okay so he did say what the axes were.
S [09:14]: Okay, yeah.
S [09:41]: That was kind of a waste of time, like, showing the whole video.
S [11:07]: It sounds like they didn't like know what they were talking about—or like maybe they knew but like didn't practice it before they were just, kinda like, recording themselves. So, like there's like a lot of like awkward pauses where, like, they're not really sure what word to say.
S [11:23]: [inaudible]
S [12:04]: [muttering under breath]
S [13:34]: Okay. Well, I guess that was it. So now do you want me to say what I would grade it as?
I [13:42]: Yes, if you could please do that, and also if you could assign a specific rating to each rubric item.
S [13:47]: Okay. So for the organization structure, like what I said before, I would give that poor cuz... I—wasn't necessarily lost, because I knew what they were talking about, but it could have been a lot easier if it was like, more well organized and they like took more time on their video and like I said, there's like no introduction, no proof—no presentation of the problem, or like result, or like previews. So I'd give that a poor. Um, they didn't like show any models. Like, any... like stick-figure drawings or anything.
S [14:16]: They didn't really discuss Newton's 2nd law very much, like, they explained a little bit in the code. Um... but I wouldn't say it was really a discussion, it was just like them saying like what was going on. Um, so I wouldn't say there was a lot of discussion there. Um, or like connections—I'd kind of give that—but they didn't mention—so I'd probably give that a fair. Um... let's see. The data—they showed us the data.
S [14:50]: Oh, they showed us the initial conditions, they like showed it um... they discussed how the parameters were adjusted... mmm. I mean they discuss the computational model, like, okay... So... I'd probably give that like a good.
S [15:13]: Um... I think they answered the what if question in the beginning. Um, so I'd give that a good. And—the delivery was pretty good, I didn't want to—like, there were pauses and whatever—so I'd give that a good, too. I dunno I might give the... I might give the content overall a fair. I don't know. It's like between a fair and a good. But yeah.
I [15:47]: Okay, excellent. So I am now sending you the expert response to that same video. I'd like you to go over it and give me your reaction. Okay, it's sent.
S [16:13]: Okay. Well the—yeah, that's fine. The fair—they did say the purpose, I just don't really like it when it's not signposted at all. Um... I mean, yeah like I said there's definitely no explanation, but they did say like "oh this is Newton's 2nd law", that's why I gave it a fair.
S [16:37]: Um... I think I said good for that one, too. Yeah. That not always clearly expressed, I definitely agree with that. Like, they sounded like they might know what they were talking about, but like, I dunno. Yeah and I gave them a fair or a good on there. On this, because... I didn't like how they like kept pausing... I don't like it when it's like walls of code or whatever. So yeah. I wouldn't have probably given them that very good, but the audio was good quality and stuff.
I [17:15]: Okay, very good. Now if you could switch the camera so I can see your face, um... I know it's hard—the light is weird in my aparment, I know it's hard to see my—okay. So let's have a little discussion about that experience you just had. How did that video rating compare to um... the rating experience you had at the beginning of the semester?
S [17:39]: Um... I think at the beginning of the semester... I dunno, it's like awkward to rate people's video when you just did the same thing. Like, I just felt bad. But like,

through the semester I got like more comfortable with it, and it didn't really bother me as much, like, if I knew someone definitely didn't have an intro I would give it a poor or whatever. Or like if they definitely had a big physics error, like, I wouldn't be as uncomfortable giving a poor. Like the first time I did it, I don't think I gave anybody a poor.

I [18:08]: Mm-hmm.

S [18:08]: But like, as it went on, I got more comfortable doing it.

I [18:12]: So—so what is it about needing to be comfortable that's so important for giving poor ratings?

S [18:18]: It just makes me feel bad, because like I'm sure they put work into it too, you know, like I'm not a physics professor, so I don't necessarily like—feel like I have a hundred percent like qualifications, like if I give someone who worked on something a poor. But like, then like I—just—tried to stay like truer to the rubric, and do it on that, like, not—rather, like—I don't know. Because like I didn't want anyone to give me a poor, so I didn't want to give anybody a poor.

I [18:45]: So you said—you said you're not a physics professor, you didn't have the qualifications to give a poor. Could you tell me a little bit about what you think those qualifications are?

S [18:55]: Well so it's like if you're a teacher like, then your job is to grade. And I'm just like a peer, so like... I wouldn't have any trouble like if it was just like, um, like peer feedback, but like knowing what like I say like determines their grade, it's just like, it put like a lot of pressure on me like, you know? Cuz like I didn't—it's not like they would know, so it's not necessarily I didn't want to hurt their feelings, but like—I don't know, like I don't... know exactly like what their thought process were for something they said or didn't say. Like, maybe they KNEW that but they just didn't say it, so... I don't know. I just give too much benefit of the doubt, I guess.

I [19:35]: Um... and your—your attitude toward that changed a bit throughout the semester? Would you say that's fair to say?

S [19:42]: I mean, it didn't necessarily change, but like I got... like I understood the rubric more.

I [19:49]: Mm-hmm.

S [19:49]: So like I would feel like I knew, like what to give poor for, and what not—or poor, or fair. Something that wasn't good for, like, some—if I saw that, or not.

I [20:01]: Mm-hmm. Well, so—uh, speaking of which, um, when you were filling out your evaluation for this video, there was a point when you were looking at your—you gave a good for the content overall, and then you decided to go back right at the end and downgrade it to a fair. Can you tell me a little bit about your thought process there?

S [20:21]: Um, because I—I didn't... well like, cuz I think I decided that one like right at the end, like not while I was watching the video, cuz sometimes I forget when I'm grading it, and then either like go back and watch the video, or think about it for a second. And usually like it changes my mind when I think about it.

I [20:42]: So you've been doing this—this is something you've done uh, throughout the semester, is going back at the end and rethinking some of the ratings you've given?

S [20:51]: Yeah.

I [20:52]: When you—

S [20:54]: Like I try to like when I have to go on Webassign, I try to like, type notes, like where you're supposed to type notes, and then decide the grades after.

I [21:00]: Mm-hmm.

S [21:00]: But then when like, "oh, I don't really know" or whatever, I like go back and like try to watch—rewatch the video, or like just rewatch a section.

I [21:09]: And when you do that, what sort of things are you paying attention to?

S [21:13]: Just like whatever I have a concern about. So.

I [21:17]: And when you do make your final decision, do you feel like you have a particular tendency? Do you tend to give higher ratings, or give lower ratings, or keep them the same?

S [21:28]: It just depends. Like on... what I rewatch and like rethink about, like, not really a tendency either way.

I [21:36]: Okay. Um... so a slightly different topic. So I noticed twice when you were doing your evaluation, you made a big point of pointing out, um, that the author of that video might or might not know what they were talking about, but sounded like they were just recording themselves, or not expressing themselves clearly—

S [21:57]: Yeah, I don't—I don't like it when they're like not practiced, or like I dunno. Like it just sounded like he didn't know what he was going to say next.

I [22:07]: Mm-hmm.

S [22:07]: You know? I don't like that cuz like, if he had been working on this video for a long time and like put a lot of time into it, and like you'd really done your work, then you shouldn't be sounding like that. But then, at the same time I like know that some people don't like to talk in front of a camera, so...

I [22:22]: So how do you navigate that? How do you navigate—how do you thread that needle?

S [22:29]: Um... usually, if it's a thing like that, like, if the rest of their video is like really sounds and good, then I just ignore it, but like he didn't really mention any like—he did say Newton's second law, but he didn't really like explain anything so like I just... took that out of like, he probably didn't like spend a lot of time on his video. And that's why I gave him like the lower production score.

I [22:51]: Mm-hmm. Um, and so over the course of the semes—so that was a lab one video which you just watched.

S [22:59]: Mm-hmm.

I [23:00]: In your estimation, was there an improvement in the quality of the videos, uh, as—as the semester went on? As you [talking over: got into labs two and three and four?]

S [23:08]: [talking over—Yeah, yeah, yes, like] um... even in my lab one video, like, I didn't use a powerpoint, like have it very organized. I don't think I had any intro, barely. And then like, by lab four I had the complete intro, like, things were very like signposted and organized, and I had—I made sure I went through the rubric and like checked do I have this? do I have this? do I have this? So definitely from my lab one video to like my lab four and five, like, they're much different and I think they got a lot better. And so did everybody's.

I [23:37]: And where did that improvement come from, do you think? Was it specific instruction, or... what was it?

S [23:42]: Well I don't really know. Like the first time I made a lab—I'd never made a lab video before, so first one... I dunno. I, I dunno if I just expected it to not be graded so scrupulously like they were. So like once I realized like, oh wow, I got—I got like a seventy on my first lab. Or something, I dunno. Something pretty low, so it was like, dang! Other students work harder on this, and then I just put more time into understanding the rubric, and like what people were grading on. And then my videos [invariably?] got better.

I [24:17]: Okay. And on the side of uh, not production, but evaluating the videos, um, do you feel that... uh, your evaluation practices improved over the semester?

S [24:29]: Yeah. Like my first like calibration was really low, and by the end they were like in the eighties.

I [24:37]: And do you think—do you think your peers were also having that same sort of improvement as you? Did you get that impression?

S [24:44]: I don't know. I didn't really talk to anybody that did this. And... I don't know. I got like, once I like got my lab videos figured out, I got like the same grade through. So I think it kind of stayed the same. But I'm sure like people understood more. [At that point, when we were done?]

I [25:03]: Do—did you pay a lot of attention to the um, uh, the comments, the feedback that you would get from review?

I [25:10]: Yeah. Sometimes. I read them. But they were pretty basic, like... "oh, I liked your powerpoint" or "she did mention newton's second law" or something.

S [25:25]: They weren't like too, like helpful.

I [25:27]: Uh, can you tell me a little bit about the hangout sessions, uh, where, uh, that were related to the labs?

S [25:37]: Yeah, I mean I liked those. I thought they were more helpful towards the beginning, and then toward the end, I don't know. They weren't helpful because like, it got to the point where like it was getting pretty difficult, and like no one really understood what they were talking about. Like, one hundred percent. And then my TA was like juggling two groups so he wasn't necessarily always there to like, help us, and we like had to wait or whatever, and then we wouldn't be able to finish all the problems cuz, like the two hours would be up. So that was not that helpful toward the end, but towards the beginning I thought they were like, super helpful, and like taught me a lot. And I really like my TA, I thought he was a good TA.

S [26:21]: And like, working with one of my other peers, Dave, he was super helpful. And like just having bouncing ideas off each other, trying to figure out if that was... it was... I liked it.

I [26:32]: Yeah, tell me a little bit more about your interactions with Dave.

S [26:37]: I mean, we would be just like, "oh, do you think you'd solve it this way?" and he'd be like "oh, I don't know, let's try, what'd you get?" "I got this" "I got something—" or "no, we didn't get the same thing, then it's probably not right, let's try a different way" "do you think this would work?" "yeah let's try that" just like over and over again until we got like the same answer.

I [27:00]: Wonderful. So back on the topic of uh, learning to—or, you know, developing your practices in evaluating these videos, uh, were there any—did you make use of any external sources? Uh, you specifically related to the rubric, or to peer assessment practices throughout the semester?

S [27:21]: No.

I [27:22]: Okay. So there was a document that was posted on Piazza, uh, by someone, that had um, a much more elaborated rubric on it. Did you ever see that?

S [27:34]: No I didn't.

I [27:35]: Okay. I didn't even know about it, one previous interview participant, uh, brought it up.

S [27:44]: No, I didn't see that.

I [27:45]: Okay. Um... what resources or videos or instruction do you wish you could have given your earlier self? Uh, in preparing and evaluating these lab videos.

S [28:04]: Um...like... just for the first one, I would have made myself understand the rubric more. Like, really like stick to the rubric, not just like take it as like guidelines. Like, that's how people grade it, you need to stick to it.

I [28:19]: Okay. Was that, uh, did the formal instruction in the lab exercises not make it clear that uh, about the [importance of the rubric]

S [28:30]: [you mean like the videos about]

I [28:32]: [yeah, yeah]

I [28:35]: The videos you watched, and what the TAs and instructors told you.

S [28:37]: Yeah, I mean... not really. They like taught us how to do the lab on a—like definitely on the first one, walked us through the whole thing, but not really like "you need to explicitly have an intro, and explicitly say these things". I don't really think he said that.

S [28:57]: Or in like, the lab drafts or whatever—

I [29:00]: Mm-hmm

S [29:01]: Like they would just say a basic intro, but like I think if you would have said on the lab drafts, and then not like flesh that out, like you wouldn't have had a very good introduction.

I [29:10]: I see. Um... I'd like to rewind a little bit to I think our first topic of discussion, which was developing your feeling of confidence in giving low ratings over the semester. Um, tell me a bit—you mentioned that you would feel bad giving poor ratings at the beginning. Uh, tell me about how you felt at the beginning giving low ratings, as compared to now.

S [29:43]: Well I didn't give any low ratings at the beginning. Like the lowest I would give is good.

I [29:47]: Mm-hmm.

S [29:47]: So it's like, everyone's like trying to get through Tech, like I dunno. I just feel bad if I give a poor. But like when I, I—like I said, like when I understand the rubric more, like if you didn't do what's on the rubric, then... like I don't really have a choice but to give you a poor. So...

I [30:06]: Right. Um, do you find that exercise or that attitude—sorry. Different questions. Do you think that's a good attitude to develop? A useful attitude to have?

S [30:23]: Well like, I'm in a sorority, right now it's like rush season, so you pretty much do the same thing. You're trying to decide who you want in your chapter as a good new pledge class, and like if they don't meet the values of your chapter, or like the values of your rubric, like... it would be a disservice to say that they did. So...

I [30:42]: A disservice? A disservice to what, or to whom?

S [30:46]: To like, I guess like in physics, like the learning experience or like... give you a false confidence, like—I don't know. I really only like can explain it like through sorority terms.

I [31:01]: Please!

S [31:01]: Like if you—so... if you're trying to rush a girl, and you want her in your chapter, or you're trying to get knowledge in physics, and... you're a potential new member so like, that's a girl's coming-through rush, and they can get good ratings until pref night, which would be like a test for physics. And they fail pref night—then they don't get in a sorority, and... like, that could be likened to you give a very good on your lab, like, thing, and you'd think you understood what was in the lab but then you get to the test you go like, "what? my lab was about this, but I don't really understand it."
I [31:41]: Mm-hmm. And so what—so what is... um, if you had to very—just stated as briefly as possible, what is the greater good served by being very strict in grading?
S [31:56]: I think it's just, like... you're able to see more like where you're not adequate at. So like, I don't know. If you just give a basic good for everything, you're like "oh okay, I'm good." Even if you aren't. Or if you give like poor for what they're actually poor at, then they know that they need to step it up.
I [32:22]: You seem like you're speaking from personal experience.
S [32:25]: [laughs] Well that's something like I'm really doing right now, cuz I'm rushing girls. So... I'm really thinking about rush, and like scoring the girls.
I [32:36]: Have you uh—
S [32:38]: But honestly it's just like scoring them like a lab.
I [32:40]: It is. Interesting. Is this something that you've been thinking about—or an analogy that just occured to you during this interview?
S [32:52]: Well like literally the reason I couldn't do it [ed: this interview, we had to reschedule] at first is because I was rushing. Or like, polishing, which is getting ready to rush. So like, yeah. Like, it's been my life for the past week. But I mean it is a good analogy!
I [33:05]: Mm-hmm. Oh, I agree completely. Um—wonderful. It's such a good analogy that I'd like to explore it a little bit more before going on to the next part of the interview. Um—so when you are uh, rushing girls, as you say, is there in fact a formal rubic for it?
S [33:26]: [laughs] yes! I don't really know if I'm supposed to talk about it, but yes, there's definitely—it's called an arc score. And um...
I [33:36]: Talk only as much as you're comfortable talking.
S [33:39]: Okay. I mean yeah. It's like, respect, confidence, like all the things... so if like, a girl who I didn't necessarily like but I thought she had a lot of respect for me, um, it's like up to me if I want to give her a 5 for respect cuz she really did have respect, but I didn't really think she would fit in my chapter, I could give her a five, but then she might get in the chapter when I know she wouldn't be a good fit. And she wouldn't learn anything from being in my sorority. I could do that, cuz I don't want to be mean to anyone or unfair, but then it's kind of like unfair to my sisters in my chapter as a whole because I don't want to give anyone a bad rating. But that's what it's there for, it's so we don't get anyone who doesn't fit.
I [34:26]: Hmm.
S [34:26]: When they shouldn't be. Or... we take away their chance to be in a chapter where they would learn something.
I [34:33]: Interesting. So, and how many times do you evaluate uh, sorority rushers?
S [34:43]: Um, on the first day, you talk to—three times twelve. There's five days. So that'd be thirty-six girls on day one.
I [34:52]: But does each individual girl get evaluated multiple times?
S [34:56]: Yeah. By three people. [unintelligible]
I [34:58]: Ah. By three people. But is it—is this stretched over time—so what I'm trying to bring the analogy together is, in the physics labs, the four labs happened at different points in the semester, and ideally—you know, we'd hoped that, we'd designed the course to help develop certain physics concepts and communication practices over the course of the semester, so if you get poor ratings at the beginning, presumably, you know, we can help you develop and get better ratings at the end.
S [35:29]: Right. So like yeah, that's how rush works too, if you get bad ratings at one chapter, you're going to get dropped by that chapter, and you're gonna like learn how to fit in somewhere else.
I [35:42]: So it's very important to choose the order in which you approach chapters, is that—is that true?
S [35:49]: No. [laughs]
I [35:51]: No.
S [35:51]: But I get what you're trying to say, like, um... like, it's very important to like know what you're grading on, and to not like... like, you're going to learn from getting dropped from a chapter.
I [36:07]: Mm-hmm.
S [36:07]: And you're going to be better next time. That you go to another chapter. So it's like the same thing, I got a really low grade on my lab 1, and I learned on like what to change to make my lab 2 better, and I got a way better grade for lab 2, and then I got a way better grade than I got on lab 2 on lab 3.
I [36:24]: So, uh, overall, if you had to give a brief summary about your attitudes toward, you know, um, the entire laboratory system, from lab 1 to lab 4, do you think it's a good system, did it work, did you get something meaningful out of it...?
S [36:40]: I mean...
I [36:41]: And what sort of problems did it have?
S [36:41]: Okay, well I will say, like I got a seventy from the lab report thing on my first one, and then I was like, that's really low, like whatever. And I asked my TA about it, and he was like "oh I'll regrade it". I got an 83. So I do think it gives you lower grades than it would be if a teacher was grading it. And I don't really know why that is, like, I don't know. Yeah. I don't know.
S [37:06]: But, um, I mean I do think you learn from it. Like, and you really put your best stuff out there because you know, you know they're going to grade it hard. So I do think it like made me spend a lot more ti—I would spend like, two, three, four hours on my lab report, I mean just making the actual video. So... I do think it really pushed me to make good content, and like, not rush through it, or not wait till the last minute to do it. Because I really wanted to do a good job, make a good grade.
I [37:35]: Hm. So, that's interesting, you said that on your first lab, uh, you got what seems to have been a lower grade than you would have gotten than if it had just been graded by a teacher in the first place.
S [37:47]: Mm-hmm.
I [37:47]: But when you were grading other videos in lab one, you said you didn't give any low scores at all.
S [37:51]: No.
I [37:51]: So do you think—
S [37:52]: But I know some people did, cuz I got a seventy.
I [37:54]: Right. So, so what do you think—
S [37:57]: I think a lot of people came in more, like, comfortable with like giving low scores. Which is fine.
I [38:03]: And why do you think they might be comfortable to do that?
S [38:06]: I mean, first of all, that's just how some people at Tech are, like, you know. That's just the culture of Tech. And then like secondly like maybe they did something like that before, or whatever. I never had, so...
I [38:19]: Okay, now so you've had no peer grading experience in the past—
S [38:26]: But I had like the peer evaluation where you just give comments, where you just said what you liked and what you didn't like, I just hadn't done assigning, like, a grade.
I [38:33]: Okay. Uh—
S [38:33]: Also, like, you know how when you give the grade and then you press submit and it says the grade?
I [38:39]: Mm-hmm.
S [38:41]: Like if I see that grade really low, I try to go back and [laughs] do like, "uh, did he—did I miss something anywhere where I can give like a little bit of a higher grade?"
I [38:49]: What counts as a low grade for you, in that sense?
S [38:52]: Like, below a 75.
I [38:55]: Mmm. And what—why—where are you drawing that from? Why a 75?
S [39:04]: Umm, I dunno. Just—I think 75 or above is like, I did okay. But like below 75 is like, dang. Like... probably didn't know what I was talking about, or like just really didn't try. Or...
S [39:23]: So if I see that grade but I'm like, oh, he tried, then I try to make it better. But if I can't, then I can't.
I [39:29]: Um... I think... let's see. In terms of your... expertise with physics problem-solving and the development of physics concepts, understanding—say—the raw physics, understanding Newton's second law, or the energy principle... how would you say you compare now to when you entered the class?
S [39:54]: I think it's way more. Um, I'm not like really good with like [laughs] explaining like, physical terms, but like I definitely know like, what it's about.
I [40:07]: Ah, so—
S [40:09]: [unintelligible] understand it.
I [40:12]: Ah, so you're making a distinction between knowing about a topic, a physics concept, and being able to explain it well.
S [40:18]: I mean I think I can explain it well, but like, only to people who like understand I explain things. Like I don't think I could lecture about it.
I [40:27]: How do you explain things?
S [40:27]: [laughs] I dunno, like very basically, I guess. Like... [laughs] with sorority analogies. So..
I [40:37]: That was a fine analogy by the way.
S [40:41]: Thank you.
I [40:43]: So when you are making these videos, what sort of audience are you making them for? What sort of knowledge do you assume on the part of the audience?
S [40:52]: I try to really assume no knowledge, because that's kind of what the rubric says. Like you need to explicitly state what you're trying to say.
I [41:01]: And do—
S [41:02]: So I really try to start from the bottom, and just, say, like, explain everything.
I [41:08]: And did you always do that, starting from Lab 1?
S [41:09]: No. I didn't—I don't think I did that at all starting at Lab 1.
I [41:15]: So what audience were you making it for at Lab 1?
S [41:15]: People who, like did Lab 1. Like they should know what I'm talking about.
I [41:19]: Gotcha. Um, and, what convinced you to change your target audience from your peers who also did the lab to the layperson off the street?
S [41:31]: Just re-reading the rubric, and like understand—like I think a few comments on my Lab 1 said "you need to explain more" or "you need to flesh this out more". So that's what I did.
I [41:42]: Alright, very cool. Um, Are there any other matters related to this discussion that we just had that you'd like to bring up?
S [41:52]: No, I don't think so.
I [41:53]: Okay, in that case let's m ove onto the next part of the interview. If I could see your screen again, and if you could open up the ball rolling example file.
S [42:03]: Okay.
I [42:04]: And you saw this on our first interview, oh those many weeks ago, um, it's ball rolling across the screen, and then a tracker screenshot of um, trajectory information about that ball. And what I'd like you to do is watch the video, examine, watch the tracker section, and then just explain the physics to me, particularly of the um, position vs time plot that shows up in Tracker.
S [42:32]: Okay.
S [43:15]: Okay, what do you want me to explain?
I [43:19]: In a—in a single—what I would like you explain is the position vs plot in the top right corner. What does it mean, and why does it have the shape that it does?
S [43:30]: Um, well, like if it was a model it would just have a straight like, linear shape. But this one has like, pretty much a linear shape, but you can see there's like a little curvature to it.
I [43:44]: And why would a model have the perfectly linear shape?
S [43:47]: Because um, like it's moving at a constant rate. Like, no... forces on it that like—no net forces on it, so that it would accelerate or decelerate or anything like that. So this is just like, supposedly moving like, at a constant speed, um, but it's still um moving. And then... yeah. Like, I don't know, it just travels through—but you can see it's a little bit curved, so maybe there's like some bumps on the ground that made it accelerate or decelerate, or made it not go completely like in the constant speed.
I [44:26]: So tell me a little bit about that, as it relates to, uh, the net force on the ball.
S [44:32]: Well like, if there was like bumps, and it like—was bouncing a little bit or something, then that would have made a positive or negative net force, and not a zero net force. So it would have—if there is a positive or negative net force, that means there's not a constant speed.
I [44:49]: And overall, would you say there's a positive or negative net force acting on the ball, as indicated by the trajectory.
S [44:56]: Um... I think, like, positive, beause it looks like it curves up a little bit.
I [45:05]: Okay. And what could the origin of this force be?

S [45:08]: Um, just like, friction, on the ground.
I [45:11]: Mm-hmm.
I [45:13]: And by positive net force, which direction do you mean that it's pointing in.
S [45:19]: It's like pointing in like, the left.
I [45:22]: Okay. So it's pointing against the direction of motion.
S [45:27]: Yeah.
I [45:29]: In what sense is that a positive net force?
S [45:31]: I don't know. I just said pos—I don't know. I just said positive because it would be—like it would slow the ball down [to the left?]
I [45:39]: Okay, okay. Alright. Um, and so, so you say the friction might be slowing the ball down.
S [45:48]: Yeah.
I [45:49]: Uh, tell me a little bit more about that. How could we eliminate system from the system, and what would look different if we did?
S [45:56]: Well, if we like eliminated anything that would give a force, then it would be a completely straight, like, line—like straight, not just straight, but linear line.
I [46:08]: Mm-hmm. And what sort of—what sort of thing, like, in the actual physical system itself, what are the sort of things that are causing friction.
S [46:18]: The floor. We would need like a frictionless floor. I dunno, like maybe some ice, like perfect ice or something. I don't know. I don't know what friction is. But something made out of that.
I [46:29]: Well there are certainly things that have less friction than whatever that surface is [it looks like concrete]
S [46:34]: [like when ice or] like a skating rink or something.
I [46:40]: Mm-hmm. Mm-hmm. So describe the—this is going to be a somewhat abstract question. Um, but what we're plotting on the top right there, is a single point, the position of a single point at various periods in time.
S [47:01]: Yeah.
I [47:01]: But the ball itself of course is not a single point. Right, it's a solid object. So why is it meaningful at all to talk about the trajectory of that imaginary point, when we're really trying to describe the motion of this, you know, solid object.
S [47:20]: The little points are like each position. Which gets updated through Newton's second law. And then it creates that line.
I [47:26]: But the position of what?
S [47:29]: The ball.
I [47:30]: The ball?
S [47:31]: Yeah. Cuz that's like the position of the ball, like every—or the mass or whatever, at every second that it's moving.
I [47:41]: Mm-hmm. And is the, uh, the fact that we're clicking in the center of the ball, is that, is that relevant? Is that the only place we would—we could click?
S [47:51]: Um... I mean, yeah, I think so.
I [48:01]: Okay. Um, and... finally, uh, how would the situation be different—would be—how would the situation be different if we were, say, rolling a hockey puck of the same mass as the ball—sorry. If we were SLIDING a hockey puck across the floor, rather than rolling a tennis ball?
S [48:28]: Um, there'd probably be a lot more friction. Like on the same floor.
I [48:33]: Yeah, tell me a little bit about that.
S [48:35]: Um... like more surface contact, and like it's sliding.
I [48:40]: So—
S [48:41]: It's not rolling, like it's going to get stopped by the floor sooner rather than later. It's not going to just keep rolling.
I [48:48]: Yeah, so what does rollling have to do with friction?
S [48:52]: Mmm.. I don't...
I [48:58]: I ask because you mentioned that the, uh, the hockey puck has a larger contact area—
S [49:07]: Yeah.
I [49:09]: —and also sliding, not rolling. And I wonder what you mean by that second one, sliding not rolling.
S [49:14]: Like I feel like rolling like the floor kind of carries itself. Like I know it's just one—like I know that rolling's just like you push it once, and it rolls. But like—the circular, like motion, that it can take itself over, I feel like it's less affected by friction than just like a sliding hockey puck would be.
I [49:36]: Hm. Could you elaborate on that for a bit? Because it's an interesting idea, that if something is rolling it's less affected by friction.
S [49:42]: Um... I mean I know that's like not really true, but like, I don't know. I just feel like if you have a perfectly circular object, then not a circular object, not rolling, like... balls just roll. [laughs] Like, you know, and hockey pucks don't roll. They just slide.
I [50:04]: So, let's go back to the very beginning of that video, before it was put into tracker, when you could still see the hand as it was pushing the ball.
I [50:20]: Right, so, I want you to imagine what would happen, um, if that surface were perfectly frictionless. And the ball received the same push that it did. I want you to tell me how the ball would move subsequently.
S [50:42]: Um, I think... it was perferctly frictionless, it would roll for a lot farther, then... it should just move like at a complete constant speed that didn't like, dip as soon as the hand pushed it.
I [50:58]: And likewise, what would happen if we had a perfectly frictionless surface, and that same hand with the same push pushed a hockey puck.
S [51:08]: It would do the, like, it would go the same. If it was the same mass, it should go like the same. Like...
I [51:17]: I want to explore again... you had what I think is a statement worth diving into a bit, that you know, the hockey puck just slides, the ball just rolls. What is it about balls that roll? That make them—that make—why are those two concepts so closely associated?
S [51:45]: I don't—they just do! [laughs] They're just circular—roll, like wheels, I guess.
I [51:54]: Um, and so—so if, if the—rolling is rotation, right, the ball is rotating as it moves.
S [52:03]: Right. Mm-hmm.
I [52:04]: Tell me a little bit about the dynamics of rotation and how, you know, in terms of forces and so forth, how you make an object begin to rotate.
S [52:17]: You just apply a force to like—a rotating object, and then it begins to rotate.
I [52:23]: But what if the object isn't already rotating?
S [52:28]: Um... I don't know. I don't really know—I don't know what you're talking about [laughs]
I [52:37]: Did you co—in this course, did you cover, um, angular momentum?
S [52:40]: Yeah. Okay... yeah, I mean... yeah but like—how to make it rotate, you just push it.
I [52:50]: Well, what's the angular momentum of a stationary, non-rotating ball?
S [52:57]: If it's stationary, there's none.
I [53:00]: And the angular momentum of a rotating ball is nonzero, right.
S [53:04]: Right.
I [53:05]: And what is... what do we can angular rotation, um... sorry. How do you change angular rotation over time?
S [53:24]: Um...
I [53:26]: You do something to the system...
S [53:27]: You apply a force to it.
I [53:28]: You apply a force to it, but with rotation, it's a particular kind of force, um, and it's really—it doesn't have the units of force, it's something that's analagous to a force, uh, but in the angular world it's a different entity.
S [53:45]: I don't know.
I [53:46]: Okay. That's alright. The—I—was looking to explore a bit into torque.
S [53:52]: Ohhh! Okay, yeah. I wasn't too good on that one.
I [53:58]: And so—to—so just briefly what you know, or, tell me what you can tell me about torque.
S [54:06]: Um... it's just like—well there was one like question on the test, it was the last question really. You have to figure out the torque of a toy, and so, it's just like—I don't know, I didn't look at my notes before this, I didn't know it would be a test [laughs]
I [54:24]: It's, it's—it's an interview, I really just honestly want to know what you think—
S [54:27]: I know! I'm just kidding, but, like... I don't know, I really don't know the words to explain it. Like I do know what it is, and like, I would know how to figure out that problem again. I don't really know the [right words]
I [54:46]: [And that's an interesting] point you're bringing up, which is something that we talked about a little bit earlier, the gap between understanding something, or feeling that you really understand something, and being able to communicate it clearly, or feeling like you're able to communicate clearly.
S [55:01]: Yeah.
I [55:02]: And so it seems very much like you're saying that it's possible to have the one without the other.
S [55:07]: Yeah.
I [55:08]: And presumably—presumably, understanding has to come before clear communication.
S [55:15]: Yeah.
I [55:17]: So what is it that—how does one develop... or, how do you develop the clear communication?
S [55:25]: Um, well like, for me to like be able to like, be confident in like what I'm answering a question about, I really feel like I need to completely understand like everything about it. Like I definitely don't understand everything about torque, or like something. So like I dunno, I just can't really ever find the words to like explain it.
I [55:50]: "Can't find the words to explain it", do you mean in terms of vocabulary?
S [55:53]: Or like, yeah, or like definition.
I [55:56]: Mm-hmm.
S [55:57]: Like I can honestly see the definition, of like, or not really the definition, but I can see how to calculate torque from like a little graph in my mind, but... I couldn't tell you in words how to do it.
I [56:10]: Okay. Well this was a fantastic interview, thank you for, uh, volunterring your time here. We're at five o'clock now, so that's our hour done. If you—you can go ahead and switch me back to your face.
I [56:30]: There you are. So the last thing that we have to do is to arrange for your payment, it'll be forty dollars for two hours.
[etc., logistics]

F.10 Student C, Summer 2017 (Pre)

I [00]: Alright, okay, so thank you very much for agreeing to participate in this interview. Um, I'll go over what our research goals are, and what today's study is going to be like. So our research goals are to understand... um, how peer assessment works in the classroom, and how students like participate in peer assessment, the attitudes and practices that you form. And we're particularly interested in how those attitudes and practices change over the course of the semester, which is why we're interviewing you once at the beginning of the semester, and then once at the end of the semester.
I [36]: Um, so today's session is giong to be basically identical to the end of semester session. We'll start out with a warm-up physics problem, uh, where we'll introduce what's called the think-aloud protocol. So one of the things I'm going to be asking you to do throughout this interview is to... think aloud, is to narrate the thoughts that are going through your head as you're solving physics programs—uh, problems, or making decisions. During the think-aloud sections of the interview, I won't be asking you questions, and I won't really be prompting you to do anything. The only thing I'll do is that if you are quiet for a little too long, I'll prompt you to keep talking.
S [01:17]: Okay.
I [01:18]: So the first part of the interview is going to be a warm-up question where we practice the think-aloud protocol. The second part of the interview is going to be a video evaluation. I emailed you earlier uh, a, an example lab video. Uh, it's the one that begins with the Tracker image. And I'd like you to watch it and evaluate it according to the ordinary rubric just like you did during your lab evaluation exercises last week. Uh, then we'll have a discussion about that, uh, and your attitudes towards peer assessment and why you made the decisions you did for that video. And then the final part of the interview will be uh, that much shorter video that I also emailed you, of a ball rolling across the screen. Uh, and I'll ask—we'll have a discussion about the physics underlying that phenomenon. Uh, and I'll ask you to make a good—make your best shot at an explanation of the physics of the ball rolling. Much like you did for your lab 1 video.
I [02:14]: Um, and so with that, we'll begin with the think-aloud warm-up.

S [02:18]: Alright.
I [02:18]: So this is a two-part estimation question. What is the volume of the room that you're sitting in right now, and if you were to turn on an ordinary kitchen faucet, about how long would it take for that room to fill up completely with water?
S [02:40]: Um... so the volume of a room right now is uh... very... squarelike shape. So the distance between each length would be the same. So if I just cube that volume I'd get the volume.
I [02:58]: Could you do that?
S [03:00]: Yes. I'd just take a measuring stick and jsut measure each of the side lengths, and luckily the way my room was built, was the same length, width and height, so I'd just do that, using one length measurement. And... what was the second question again?
I [03:22]: The second question... sorry. The second question was if you were to turn on a kitchen faucet in that room, like an ordinary water faucet, how long would it take for that room to fill—for that volume to fill up with water.
S [03:39]: Um... I would have to measure the rate how fast the water droplet flows down. And I'd measure a single water droplet volume. And then... after I found out how fast the rate at which it falls, then I can... uh, do some calculations... where uh, I take the volume of the room I found, and then divide it with the, um, divide it with th evolume of the water droplet. And I'd find, how it'd fill up the room.
I [04:18]: Okay. Um, so that's a good—that's a good description of the strategy which you would take to uh, produce that estimation. But what I'm actually asking you to do is to make that estimation. I want you, right now, to estimate the volume of the room you're sitting in.
S [04:40]: Oh! Okay, so... I'd say, one two three, four... well... let me get out my calculator.
I [04:51]: Of course.
S [04:56]: So the volume of the room is about one hundred twenty five cubic meters. And then... if we say... that... um... a water droplet is... say like, one fluid ounce... then... I'm supposed to calculate the rate at which—or how fast it would fill up the room.
I [05:37]: Uh, so what I'm interested in is the amount of time it would take for the room to fill up. You can take whatever strategy or whatever calculation you think would produce that answer.
S [05:51]: Okay. So... well like, it... if I'm, if I'm imagining that this room was empty without any table or stand, then I'd just do 125 divided by... uh, one, if I say like the water droplet is one, and it drops one at every one seconds, then it would be 125 seconds until the room fills. But then, I know that's not the case because... I also have a bed, a lamppost, a table, so I'd have to calculate those in, too. Because it's not totally an empty room.
I [06:32]: Okay, okay, excellent. So that's actually, um... hmm... that was not quite what we were going for with the think-aloud protocol. So what, what I want to happen for the subsequent parts of the interview is to—for you to go about solving the problem as you see fit, or doing the discussion as you see fit, without directly asking me questions. I'm supposed to just be a passive observer for the think-aloud problems. So maybe let's try a different warm-up, because it is somewhat important to get this right. Thank you very much for your effort, I really do appreciate you joining in this interview.
I [07:25]: Um... I want you... I'm having to improvise a—what is a good... another estimation question that I could ask you. Um... explain to me, uh, if you would, okay.
I [08:10]: Suppose you have a pendulum clock, like a grandfather clock with a swinging pendulum. And suppose that the clock was running too fast. Every day, at noon, the clock was five minutes of where it was the previous day, at noon. What might be wrong with the clock and how might you fix it?
I [08:40]: And please talk aloud.
S [08:44]: Um, so... there might be... um... there might be some sort of mechanism that every one second, moves, to initiate the ticking time. But then if there's something wrong with that mechanism, so that instead of one tick, there's two ticks happening, then it might cause the pendulum a little bit faster than expected. So the way to fix that would be... um...
S [09:34]: I guess... measure how fast the rate of tick goes, and then, put a countermeasure so that it goes in the opposite direction to counterbalance that force. [laughs] That's my best guess.
I [09:48]: Okay, that, that was actually very good. That was the sort of self-directed explanation I was looking for. So that's how the think-aloud protocol is supposed to work. So very good. So with that, let's move on to the next part of the interview, which is the video evaluation part. So on the email that I sent you, there's a lab video and also the rubric. And I'd like you to watch the lab video, uh... and evaluate it according to the rubric, in according to whatever is the most natural manner for you, like the way you did it during your lab evaluation exercises.
S [10:33]: Um, okay. Wait, I have two videos in the...
I [10:45]: There's a larger one. Uh, there's one where the name is just a string of characters. That's the lab video that we want to watch.
S [10:54]: So, uh, ball rolling example dot mov?
I [10:59]: No, the other one.
S [11:01]: Oh, okay.
I [11:03]: Yeah.
S [11:14]: Uh, it's taking a while to download.
I [11:18]: It—it might take a moment to download. Now, while you're waiting for it to download, is the rubric legible?
S [11:30]: Yes, it is.
I [11:31]: Okay. So I do want you to evaluate this video in whatever way feels most natural to you, perhaps the way that you would—well, what I would like is the way in which you had been evaluating the videos in your previous lab exercises. So if you tend to watch the videos straight through and then do the evaluation, then do it that way. If you tend to pause and rewind, do that. Uh. But at the end I would like you to list down what ratings you give the video, and the comments that you would give the video.
S [11:58]: Okay.
I [11:59]: Mm-hmm.
I [12:42]: Alright. So let me know when the video's downloaded.
S [12:44]: Oh, uh, I've been starting.
I [12:45]: Oh! Sorry. Okay, so please—let's back up for a second. So when the video is running, uh if you could please share your screen with me. I'd like to see what's going on your screen as you're watching the video. And be sure to clear your desktop of any sensitive information before you switch over to screenshare.
S [13:08]: [laughs] okay. Uh...
S [13:22]: Wait.
S [13:29]: Is it working?
I [13:29]: Yes. Yeah, it's working.
S [13:33]: Can I...
I [13:34]: Okay. And I know that while you are watching, uh, the video and listening, it's very difficult to narr—to speak at the same time. Um, but if you have a particularly strong thought, or if you pause to collect your thoughts or write stuff down, uh, please speak aloud.
S [13:51]: Alright.
I [13:52]: Okay. Sorry for the confusion.
S [15:27]: Um. So... If I look at the organization structure, I think that there is an introduction where he introduces Newton's second law, but then, when he mentions the... coordinate system, he mentions that the axises are straight, but I think it'd be better if he mentioned how the origin was at the center of the... peanut butter cup. Or, peanut butter jar, and then... uh, he doesn't really give a... direct result. At the beginning of the problem, like a brief summary. So I should take that into account.
S [17:38]: Um. So. He is correct that... fnet, the total interactions to determine fnet is a bit difficult, because we don't exactly know which of the forces are all acting on the system. But... it—the forces are not exactly invisible. We would just say—because we do know that there is a normal force, a surface friction, perhaps air resistance. So I think that could be worded a little bit better than saying "invisible".
S [20:17]: [under breath] well... um... um.
S [20:52]: Uh, can I, can I go over my results?
I [20:55]: Of course.
S [20:57]: Okay, so, for the organize—organization structure, I said that it was good. Uh, do I give any—do I give the reasons?
I [21:09]: Yes, I would like you to explain that reason.
S [21:12]: Okay, so, there is a clear organization structure in this video. And there is an introduction, but... that introduction... does not give a preview of the major sections, where it says introduction lacks preview of major sections. But the video was easy to follow overall. So I gave a rating of good. And then for the content models, the model was clearly shown, and he does discuss that the reason behind fnet was zero is because there is constant velocity. And then...
S [21:50]: I gave that a rating of very good. And then the content prediction discussion I gave that a rating of good, because he, he does use Google Spreadsheets and then calculates a position versus time graph. But then I just felt that the reasoning was not very thorough, because he says that the observerd motion was lower than the predicted motion, but if that's due to some velocity change—we see that in this graph that in the blue line, the slope is actually fluctuating where it goes from above to downwards to above a little bit, so maybe he could explain that a bit further, whether it be human error from manually setting the position mass points in tracker, or if there's some other force into—that's taking into account. And the content overall, I said it was uh, I thought it was good becauese he does incorporate the what if question on I believe it was about the coordinate system, and another what if question about um, fnet. But he doesn't really explain why the observed motion and the what if question for fnet is so... uh, directly correlated. Which was the second question I believe on Webassign. And production delivery I said it was excellent because uh there weren't any distracting noises in the background, the lighting was very good, the audio was very easily heard, and uh, the vocal quality was constant throughout, so there wasn't an issue with that.
I [23:45]: Okay, very good. Um, so leave your screen uh, up, I'm going to email you the instructor evaluation of that same video. And I'd like you to read over that and then, uh, tell me what you think about it.
S [24:01]: Okay.
I [24:06]: Alright, I've just sent it over email.
S [24:32]: Okay. Um. Here.
I [24:37]: Yes.
S [24:40]: Um... so... wait.
S [24:48]: Um. It's not working. Okay.
S [25:04]: So the fair is for the introduction?
I [25:07]: Yes, item one two three four five.
S [25:12]: For—organization and structure. So it, uh, what do I do with this?
I [25:20]: Read over it, compare it to your ratings, and tell me what you think about it.
S [25:26]: Okay, so, the instructor answer key says fair for organization structure. He says no results in intro... no preview of major sections, uh, that's like the same thing I said. And then good description of lab purpose. Uh, WDIM and WI should have been signposted. Um, I'm not sure what this means.
I [25:51]: Mm-hmm.
S [25:53]: But, uh, I gave a rating of good—of fair. So they are slightly similar. Where, uh, the only major difference was the—I didn't—I didn't take into account WDIM and WI. The—thing that varies the most with my analysis and the instructor's analysis was that the... um, content model... the instructor gave a poor—I gave a very good ratings. So that's a clear contrast. The reason I gave a very good rating and I gave a poor rating was he said there was no explanation of specific physics principles or connection between the system. I can—I can kind of see where he's going, because now that I think about it, the way he presented the, um, vpython model was that he never connected it to Newton's second law, but rather he kind of stated the equations that were already given in vpython, uh, code. But I—I just mentally, I just kinda associated that with Newton's second law already, even though he didn't explicitly mention the connection between the code and the update of position, velocity.
I [27:27]: Mm-hmm.
S [27:29]: And then... for the content prediction discussion, the instructor gave a rating of... very good observation of model comparison with a good account of the role of friction as it relates to plot divergence as time goes on. The account of the initial condition calculation was buried in a disorganized explanation in—uh—of the python code.
S [27:56]: Uh, I never really took into account the organization structure on that. I kind of lumped that together with the, uh, the organization structure at the top. But then I—he, he does, uh... basically—but the way I thought of it was even if there was friction, the slope of the line was clearly going from negative to positive, negative to positive, so if there was friction, then why would there be a fluctuation between those slopes. That's the way I kind of thought of it, a kind of errorwise graph. But... we gave the same ratings, good and good. So the next one is content overall. Uh, the instructor gave a rating of

good. Inadequate WDIM explanation, no motion net force connection. Uh that's when I ment—I also gave a rating of good, the same as the instructor, because on the what if question on Webassign part two, I believe there was a question that asked about the net force and how well the connection—that is connected with the model. And I just felt that wasn't sufficient enough.

S [29:12]: And then... um... the quality uh, the instructor said very good, very good production quality except for hard-to-read lines of code. I gave a rating of excellent because... um, I mean I thought the code was quite visible, and there weren't any major errors uh with the code itself. And it was very audible to hear. So it's just a difference between excellent and very good.

I [29:43]: Okay, very good. Thank you, uh, thank you for that very thorough explanation. So now we're moving into a more traditional sort of interview style where I'll ask questions and you respond and we go back and forth. Um, so tell me a little bit about your experience rating this video and previous videos in your lab, and how your ratings would tend to compare with the expert ratings.

S [30:05]: Um, oftentimes it would... it would just be a difference between uh, one, one chart. Like very good or excellent. But I guess that the... my, um, what's it called? Um, content model might be a bit skewed because sometimes I'd mentally associate myself when he gives up a code line, I just kinda assume, even though he doesn't verbally explain it, that that's newton's second law, or that's the update position model. Or update position equation. So... it's, it might be a little bit biased that way, because, uh, I'm already having a mental thought of "this looks great, this looks great" even though he might not be saying it verbally. So I think that's the part where that's the most, uh, most bias might take into play. Because if we look into like visuals, the visuals I believe everyone can have a very similar, uh, similar uh rating, because it's pretty standard, like, if you can't hear the audio, then you know there's something wrong. But if you can hear the audio, then it's correct.

I [31:15]: So I'd like to go back to that discussion, uh, about the content models where you said that you were... uh, it seemed like you were evaluating the code on screen without considering too much the explanation, the verbal explanation that the author was giving.

S [31:32]: Right.

I [31:32]: Um, and you said that that biased your answers. Did you think that—did that cause a bias in a particular way? Toward the low end of the scale? Toward the high end of the rating scale?

S [31:43]: It was more–because I'm more, a very visual person, it would be kinda skewed higher if the stuff I'm seeing such as the code is correct. Which I believe the code was correct by itself, where, uh newton's second law and the update position equation was correct.

S [32:00]: But if, if the code is not correct, then I would give a lower rating, uh, a lot more lower rating if than if the audio might be correct.

I [32:11]: Okay. And so, when, uh—when you were... producing these videos, yourself, uh, do you expect other reviewers who are peers to have that same sort of attitude toward the, uh, toward the computational code?

S [32:28]: Uh, you mean the—the content model?

I [32:31]: Yeah the content models, do you expect that other peers behave in the same way that you do when you evaluate that part?

S [32:39]: I think it's likely, because we already know, like since we have already done the lab, we already know that, uh, we already know the stats that should progress. I mean as long as there's no glaring error when he shows the running of the code, where instead of like, the force of gravity going downwards it goes upwards? I think a lot of people will give a much more higher rating than a normal, instructor rating by itself.

I [33:08]: And why—and so, tell me a little more. Why do you think that is? Why do you think that the errors would tend to be the student rating is higher than the instructor rating?

S [33:18]: It's—it's somewhat difficult to pinpoint the exact, uh, like—why how, like—it's very difficult to ponpoint how much, uh, of what content uh evaluation that the producer should give into the code.

I [33:38]: Mm-hmm

S [33:40]: Like, the code might be correct by itself, but it serves no—in the explanation we would have to take off points for that. But that—uh, evaluation of how much we should uh, consider... like if you only give—if he only gives like one explanation instead of like the necessary five, we have to, there's no clear, uh, grading system for how much [gestures] evaluation. If that makes any sense.

I [34:12]: That does make sense. It is... uh, and so when there is no clear entry on the rubric that would seem to correspond to the actual state of the video, you're required to make a judgment call.

S [34:28]: Right.

I [34:29]: So tell me about—is that—tell me about your feelings about having to make a judgment call in kind of an ambiguous situation.

S [34:38]: Usually I just try to do... give a more... uh, like... it's like kind of like a innocent until proven kind of thought. Like I start them off with a hundred, and then every time I see a glaring error, or a very, uh, eyecatching error, I kind of dock points off from there.

I [34:59]: And why, and why do you do it that way?

S [35:03]: Um. Ehh... it's just how I've always been doing it.

I [35:07]: Have you done peer evaluation in other courses, or at other institutions?

S [35:12]: Yes. In like—at Georgia Tech in English, for example. Peer review and stuff like that. In English courses.

I [35:20]: Mm-hmm. Um, and that's, was that attitude of innocent proven guilty, was that the explicity instruction that you would get during your English courses, or was that something that you just felt was the right thing to do, or that you decided for some reason to do?

S [35:36]: I just felt that it was like, the right way to approach it, instead of, like... instead of like—because it might be difficult to pinpoint all the errors, but it's very easy to pinpoint some of the errors, and if I look at it that way, then it would be a more positive note to the other person, too.

I [35:56]: And do you think the experts here who are assigning the grades in this physics course, do you think they take the same approach?

S [36:03]: Um... I'm not sure about that.

I [36:07]: What—yeah. Tell me what you think they might be doing.

S [36:11]: Uh, I believe like, since, uh, I—I... I'm not sure if they start up with 100 and then dock points off, but... I mean that could be the case, I'm not sure. But they could also just be, like, uh, looking for major errors and then kind of base it off there. Without exactly like starting off from 100. But, uh, I'm not sure about that.

I [36:36]: Would—how would you... how would you find out, if you weren't sure about that? And do you think it's a—I would like to expand this idea, uh, a bit, about the distinction between the way you approach rating, and the way that you think your peers might approach rating, and the way that you think that the experts might approach the rating. Um, so, one particular question is, do you think that you are uh, typical. Do you think that your way of assigning ratings uh, is typical of the other students in your course here?

S [37:14]: Uh, I... um... yes? But that's just... because based on my previous, like, peer evaluations in some other course, like English, that's just from my experience in those classes, that they are pretty similar in terms of rating.

I [37:36]: Mm-hmm. Um... do you think—I'm going to use a provocative term here, do you think the instructors are being harsher than you? Or you are being nicer, uh, than the instructors?

S [37:49]: Uh, I think, I probably think the instructors have a more harsher grading system, but I don't think that's necessarily a bad thing. It could be perhaps that I kind of missed an—a crucial aspect of the, of the, uh, evaluation. Because the instructors, like, they obviously know their material well. So... yeah.

I [38:14]: And that—and what does knowing the material well enable the instructors to do, that you can't do?

S [38:22]: Well, like... it's like... um... they have a more thorough understanding of the material. So with that understanding of the material, you're able to analyze a lot more deeper into questions, rather than like, "oh, here's the equation, it looks right, but what's the reason behind that equation?" That kind of idea.

I [38:47]: Mm-hmmm. Okay. Very cool. Um... so I might have some follow-up questions for this part at the very end of our interview, but for now, let's move on to the last part of our session today, which is that small tennis ball video, the ball rolling example, that I emailed you. So if you could please switch over to your screen, um, as—share your screen with me.

S [39:14]: Uh, is that ball rolling example dot mov?

I [39:17]: Yes.

S [39:18]: Alright. Um... do I, uh, share the screen?

I [39:29]: Yes, please.

I [39:38]: Okay, and so what I'd like you to do is, uh, watch the video, um, and explain to me the physics underlying this situation. You'll see the ball rolling, and then you'll see a tracker analysis afterward. And I would like just your explanation for why we found the results that we did.

S [39:57]: But this is, uh, different than the peanut butter roll...

I [40:01]: That is correct. This is a raw observational video. And so I want you to come up with your own explanation of the phyics underlying it.

S [40:09]: Okay.

I [40:10]: Yeah. And you could think aloud here—tell me what you're noticing and thinking.

S [40:16]: Um, so, we notice that...

S [40:55]: The position versus time graph is almost a straight line [inaudible], so we do see a slight curvature at the very beginning. And I believe that's because... uh, lemme check. [seeks back]

S [41:11]: Yeah. It's... he gives a initial velocity at the very beginning, so the velocity is changing between a couple—these couple frames right here, where he starts to touch the ball, and then move it forward. But near... near the end, because it's just moving—it's just moving like this, we see that at this portion, the velocity is constant. And if we look at the Tracker position vs time again... yeah, it's all—the—uh, the curve is a lot straighter near the end than near the beginning from zero to point two seconds.

I [41:53]: So can you go back and explain to me... um, show me the point in the Tracker data where the push, the initial velocity is causing the curvature?

S [42:05]: Uh, yeah. [seeks to beginning of video] Right here.

I [42:11]: Okay. And now show me the corresponding moment on the Tracker data.

S [42:16]: Um...

I [42:17]: Yeah, the, just take me to the part of the video...

S [42:21]: Obviously it would be right here. [points to t=0 on p v t plot]

I [42:24]: It'd be right there. Now is that... um, show me the position of—point zero there, that's the first point in the Tracker. Is the hand actually in contact with the ball at that point? Cuz look, look where the trajectory of the ball starts, look where the origin is.

S [42:42]: The origin is... [points to origin]

I [42:45]: Is there. Yeah. And so what's the hand doing when the ball is at that point?

S [42:52]: Um...

S [42:59]: Uh there's no hand.

I [43:09]: So can you tell me a little more about...

S [43:12]: Oh, I see?

I [43:13]: Yeah, so what's causing the curvature now, yeah.

S [43:14]: It might be, um, that constant velocity is still not reached yet because there is still a bit of changing in, uh, velocity. But eventually it reaches constant velocity. Because the force that the hand pushes is still acting these few frames, which we see a slight curvature.

I [43:38]: Okay. Um, are there—

S [43:41]: But there—

I [43:43]: Okay.

S [43:43]: Wait. Wait, what?

I [43:46]: Are there other, um, are there other things going on that might also contribute to the curvature?

S [43:53]: Yes. We also see the ground which also figures, some frictional force. So... it can't—it can't always be a straight line, like it's very close because, uh, I believe that's a tennis ball. And the ground is not, it's not some, um... something that causes a lot of friction like, say, a carpet. So it's not very... um... extreme to the point where we see a very big curvature. But we do see a slight uh, [inaudible] we construct a computational model and we see our experimental data, this uh, the computational model will overpredict because it does not account for the outside forces.

I [44:41]: Okay, the computational model will not account for the outside forces. So could you just enumerate—could you list those, uh, outside forces that you think are in play here?

S [44:52]: Um... maybe a very slight air resistance. Very small air resistance force. And also we would have the frictional force.
I [45:03]: Mm-hmm.
S [45:04]: Near the end of the, um... ball rolling.
I [45:09]: Uh, and how might we reduce uh, air resistance and how might we reduce, uh, the frictional force?
S [45:18]: Uh, I would say air—because, uh, the tennis ball is a very small uh mass, air resistance would not be a very extreme uh, extreme like cause [inaudible], but in order to reduce the effects of friction, we could have a more... uh, like smoother surface, uh, floor. Say like a bowling ball in a bowling alley. If we roll a bowling ball in a bowling alley. If we use tracker to evaluate that, I believe the... um, line would be less curved.
I [46:01]: Okay, um... very good. Thank you very much. Uh so I think that's going to be end of our interview itself, I have one followup question. Uh, could you tell me the name of the course, uh, that you took, um, for English instruction at Tech. I'd like to examine how the peer evaluation system worked in that course.
S [46:25]: Wait. Could you repeat that one more time.
I [46:27]: Yes, sorry. Could you tell me what English course you took at Tech? I would like, I would like to research how the peer evaluation system worked.
S [46:35]: English, uh eleven oh two.
I [46:35]: 1102. Okay, very cool. Uh, and peer evaluation was uh, can you tell me how it actually worked in that course?
S [46:35]: Um, so, I believe we were paired up with two, with two... with uh, two, one or two other people. And then, uh, we were given a piece of paper where we were to list our thoughts about the um, critical analysis of the essay.
I [47:09]: Mm-hmm.
S [47:10]: And then if we saw any major grammatical errors, or a lack of organization structure as we flowed through the paper, we would just give a... instead of like the ratings, we would give, uh, an essay explanation, kind of. Like you wouldn't be good, bad, excellent, it would be like—what did you think of this as a... written explanation. Instead of a grade, per se.
I [47:36]: Interesting, okay. And you'd be—but you'd be paired up, and you would know who the other person was.
S [47:42]: Yes.
I [47:43]: Right, unlike this where it's anonymous.
S [47:46]: Yes.
I [47:47]: Okay. Um... and is—were you required to take this English course?
S [47:53]: Uh, yeah.
I [47:55]: Uh, is—is—is English your second language?
S [48:01]: Uh no, it's first language. I mean—like—all of Georgia Tech students have to take English 1101 and 1102, but—
I [48:11]: Oh, okay, I—
S [48:11]: —I had AP credits for 1101, so I just went straight to 1102.
I [48:13]: Okay, I was confused, because if you had said 1101, I would have known what was going on. I thought 1102 was the required course for English language learners. Okay, I was slightly confused.
[logistics etc.]

F.11 Student C, Summer 2017 (Post)

[logistics, etc.]
I [07:37]: And remember, while you're watching the video, it's not strictly necessary for you to, uh, speak aloud everything you're thinking, because I know that's very difficult to do while you're watching and listening to something else. Uh. But if you pause it or rewind it, or a particularly significant thought occurs to you, please speak it aloud.
S [07:55]: Alright, cool.
S [08:18]: Wait, what was the...
S [09:15]: Alright, so the first what if question was answered. Uh...
S [09:46]: So, it's already been around one thirty seconds into the video, and I think can already start looking at the organization structure on this video format. And looking at the rubric, I can—see that there is a, a clear introduction where he states the purpose of this lab, which is to support Newton's second law. But, uh, uh, having graded other previous videos, in the course, uh, one of the most important things that I realized that is lacking in this video is that... uh, in other videos there's usually some sort of preview of other sections, like, "I'm going to show the code, and then show the results" or something like that. And there's no, uh, major preview of sections in the introduction, and uh there's no statement of result that is also included in the introduction.
S [12:11]: Um, so, uh... I think that was part of the what if question for lab one, but... uh, I wouldn't say the forces are invisible. I'd just, they they are hard to determine because they're uh, you don't really exactly know all of the forces in the surroundings. But I would just word that a bit differently instead of saying it's invisible.
S [13:16]: Okay so, uh, in—having seen this part of the video, I can, I can now evaluate the content models. So, um, he does identify Newton's second law in one line of code, but I do not think that he fully discussed the main specific principles in the model because... uh... he never really went into depth into Newton's second law. And uh, looking back I can now uh, see that his purpose was kind of a bit too vague. He might have—he could have included something about constant velocity, so we can see what about Newton's second law we could be evaluating.
S [15:31]: Um. It should be "lower", but "slower" works. Like, he gets his point accross. So... wait, that's the end of the video.
S [15:47]: Alright. So, do you want me to... say my grading scale?
I [15:52]: Yes, please.
S [15:54]: Alright. So. For organization structure, I'll just say it's good, because he does, uh state his purpose very clearly, which is to explore Newton's second law. But, um, he doesn't really go into depth about what about Newton's second law. Like, say, evaluating what happens if velocity is just constant instead of accelerating. Or when Fnet is zero. He doesn't really specify the, uh, key features. And uh, I wouldn't give it a very good or excellent because he does not have a statement of results, or a preview of major sections. So I would just give it a good, for organization structure.
S [16:38]: And for content models, he does, he doesn't, he does say Newton's second law, mainly from the introduction but there's not discussion of the... physics concepts. Uh, primarily Newton's second law in the code. Which is where, uh I belive where we're supposed to look at the evaluation. So I would give that a poor.
S [17:02]: And then... for content prediction discussion, uh, he does show this data used to initialize the model which is in uh, csv file. And he does eventually convert it into google spreadsheet. And he does discuss why the observed motion is different from the predicted motion, which is... uh... due to outside forces, not accounted for. And that the observed motion which just used—or the predicted motion just used fnet equal zero. And he oes say that, uh, the data like, fits the model pretty well. So, I would give that a... very good. For content prediction discussion. And then for content overall, I would give it a... um... I can't remember, I can't really remember what the what if questions were, but one of them was about the axis, which was answered correctly. And the, um, the second one I believe was... um, something about... fnet. Or how fnet would affect—how different fnets would affect the model. And, uh, that kind of, that kind of discussion was... a bit too, uh, vague, where he said that the fnet was invisible. So I would just give content overall to be good.
S [18:30]: Because they—the questions were answered, but not, uh, the second part was not answered fully. And production delivery, um, so... I—it was—the—lighting and resolution was really good, and the audio was, uh, very easy to hear. There were not distraction, distractions, and... the production quality, uh, it was very easily... uh, everything could be very easily seen and read. And... the vocal qualities were good. So... uh, there was one discussino on Piazza where um, someone asked if making the code in just like one slide, like this, and just scrolling would be detrimental to production delivery. And, uh, I wasn't so sure about this earlier, but um, one of the instructors, uh Ed Greco, he said that these—the code should be presented largely and uh, it would be helpful if it was in a powerpoint slide. So... I think I would just dock one point off. So Make it into a very good instead of an excellent. But other than that, the production delivery was very good.
I [19:48]: Excellent. So now I am sending you the expert ratings for that same video. And I'd like you to read them and react to them.
S [19:57]: Okay.
S [20:14]: So, the expert says... fair, no results in intro. No preview of major sections. Good description [sound from interviewer's computer]... uh, sorry?
I [20:26]: Sorry. Continue.
S [20:28]: Uh, so it's a good description of lab purpose. Um... yeah I agree with that. Except I did give them a good. Uh, I'm not really sure what, um, wdim and wi is. But... yeah, it's just uh, one grading scale off.
S [20:50]: Uh... content models, uh, poor no explanation of physics principles or connection. I do agree with that, I also gave it a poor because he never really linked Newton's second law with the code itself. And... um, constant prediction discussion... very good observation model comparison, with a good account of the role of friction as it relates to plot divergence. Uh, account of initial condition calculation was buried in a disorganized explanation of the python code.
S [21:25]: So... I think... uh, I'm not sure if I gave it a good or a very good, but I do agree that, um, overall he does discuss the, um, model comparison using the google spreadsheets, which was very good. And... uh, the initial—conditions calculation was buried in a disorganized explanation of the Python code.
S [21:56]: Um, I wouldn't really—I wouldn't really uh, count that as content—content prediction discussion, I would—if it was disorganized, I would put that in the organization structure and dock points off from there, instead of including it in the content prediction discussion. Because as long as he includes it I would just count that as correct.
S [22:17]: [garbled] good. The second what if question was about net force, and the way he—he just kind of discussed things as invisible, uh... the surroundings, due to the surroundings and that you don't know the surroindings' total forces involved, it is difficult to find the total net force, but it could have been worded a bit better. Uh, and production delivery overall uh, it was, uh, very good. Yeah. I do agree.
I [22:52]: Okay! Very good. so, now we'll go into the next phase of our discussion. And if you could turn your camera to be back on your face.
S [23:01]: Um...
I [23:02]: Yeah.
S [23:08]: Uh.
I [23:11]: Excellent. And so now I'd like to talk about this later experience and the other rating experiences you've had throughout the semester. Uh, so to begin with, how did it feel, generally, to rate this video, um, at the end of the semester versus when you were just starting out your peer assessment at the beginning.
S [23:32]: Uh. I felt that I have a more knowledge into grading the scale. Because beginning on, I didn't have a lot of experience. I probably looked at uh, maybe four or three videos before going directly to grading the actual peer reviews. But now, uh, having... I, I've looked at like several, maybe about thirty videos, or like twenty-ish videos, so I kind of see a pattern into how the instructor grades the rubric. At the very beginning I just kind of look over the entire video, and just kind of look at the rubric by itself. But now I kind of notice that there's a, uh several distinct... several distinct like sections in a video that you should—you kind of break up and then look by itself.
S [24:24]: So like first it's the introduction. Once the introduction is done, just stop the video and evaluate. And then next is the content prediction model. So I just look at the model and see if the physics content's related to the code. And then if it doesn't then you can just dock points off, or just do something in that section.
S [24:43]: So I think the most important part of it is looking at the video in sections, rather than just the whole thing at one time. And that helped my... uh, calibration grade improve a lot.
I [24:54]: So do you feel like that, uh, resembles the way that experts evaluate these videos? [Like breaking them up into sections]
S [25:01]: [Yeah I do] because... I'm sorry?
I [25:04]: By breaking them up into sections, I mean.
S [25:08]: Yeah I think so. Because, like, previously I wasn't doing too well like, doing the instructor video and then my, my peer review. Like, there was a... on webassign there's something called like calibration grading.
I [25:21]: Mm-hmm.
S [25:23]: And it determines how well your responses matches the instructor's ones. And at the very beginning I wasn't doing so well, but at the very end like, near lab four lab

five, I was getting nineties and one hundreds, just basing off, uh, the new method that I tried.
I [25:39]: Interesting. And so did you—do you feel like you discovered this method, you know, naturally on your own, or naturally while working with other people? Where—what made you feel like that was the right thing to do?
S [25:52]: Um, like, well—no reason that I noticed taking this class, so... so I just kind of figured it out by myself. It was mainly just like, kind of looking after—after I grade the instructors video, There's like a little thing that pops up at the below—like at the very bottom, where it says like, how the instructors thought on these responses. And so I just like, break it up into introduction, content, and then like, content prediction. It's just, uh, worked out that way. It was just very easy to look at the instructors response and see uh, what he thought. Or she thought.
I [26:33]: So, you mentioned before that um, people were talking on Piazza about how to assign certain ratings under certain conditions. What things would make you get a very good or an excellent on, uh, the last rubric item, for instance. Were you involved in any of these discussions directly?
S [26:54]: Uh... like one of them? Uh, I posted on one of the comments. Uh, I aid that, um, so, it was about—it was very specific on the code. Um, python code display. Like how you should display the python code. Whether it should just be one window and scrolling down. Or whether it should be in a powerpoint slide, very broken up and in very big, uh, pictures. Kind of.
I [27:21]: Mm-hmm.
S [27:22]: Like a very big font. So... when I did my first lab report, uh, I just made the python code into one window, ad was just scrolling down showing what happened. Instead of having a powerpoint slide breaking up the code. And uh, one of the peer—peer uh—peer students, uh they, they docked points off saying that the code was a bit too hard to read. So, I just, uh gave out my personal experience to uh piazza and said, uh, it might be a little bit better if you write—you break up the code, and then make it into big powerpoint slides. So it's a lot easier to read. And more, uh, comprehensible I guess. Like easy to follow.
I [28:06]: Mm-hmmm.
S [28:08]: But I just gave my opinion on that in Piazza.
I [28:13]: Wonderful. Can you uh, so did you... uh so this pattern that you recognized of you getting uh, higher calibration grades when you view the video as being composed of sections. Uh, did you ever communicate that to any of your classmates?
S [28:32]: Uh, no not really [laughs] There was—yeah I didn't really know anyone in the class.
I [28:39]: What about during your—what about your hangouts with your TAs, your small group hangouts. Would you ever—would y'all ever talk about peer grades—peer rating strategies?
S [28:50]: No. Not at all [laughs]
I [28:52]: What would you do in those sessions?
S [28:54]: Um, so there's uh, two sessions, one um, we can all decide on times, but I choose mine to be tuesday and thursday, and on the tuesday ones we would have a group problem solving session, where, um, usually we were given about four... four five or three problems. And we would just all work on it together.
I [29:17]: Mm-hmm.
S [29:18]: So we would just be discussing how to solve the problems in that session. And then, on the thursday session we would have lab drafts, and uh we wouldn't... like, the way we—the way we did lab drafts, it was just like... um, so like, it wasn't like an entirely, uh grading lab drafts. It was more like what I could improve on. So... to make it better for the final submission. So it wasn't exactly uh the same as the webassign lab uh, grading thing. It—it was all more different. It was more like, what to improve on.
I [29:57]: Okay. Um, so let's—let's talk a little bit more about your incresing calibration grades. So you feel that you're much, uh, well... you said that you feel like you're more capable of... agreeing with the expert rating now at the end of the semester, than at the beginning. And you, uh, attributed a lot of that agreement to um, your discovery of this pattern of breaking up the videos into sections. Uh, is there anything else that you think contributed to your, uh, improvement in your calibration scores? Anything else that happened in the course, or that changed in the way you look at things, or think about things?
S [30:38]: Um... like I would think the most important thing is um, just... looking at a lot of videos, because... like the good videos have a very clear pattern of what they're doing right, and—uh—kind of the very bad videos, they're very... they're like—they're very different from the good videos, like... you can clearly see a contrast. But, um, also for when you—when you're evaluating the content model, I would say it's definitely important to know your physics content very well.
S [31:14]: Because like, without knowing the physics content, it's hard to do anything about that [laughs]
I [31:20]: Alright, and so tell me a bit about the evolution of your, um, understanding of physics concepts. Like, so how much more uh, or more thoroughly do you think you understand physics content now?
S [31:32]: Um, yeah, I think I—well like, coming into this class, I didn't—I didn't know anything about physics, but uh, after there—after learning this course, uh, I feel... that... uh... it's always important to start with fundamental principles, like... uh Newton's second law and then go off from there. Because, uh... in high school, I took physics in I think tenth grade, and it was a bit very vague. Because I was just memorizing formulas here and there. But I really like how this course says... uh—start off with a fundamental principle, say like Newton's second law, or the work-energy principle, and then based off your understanding from those key concepts.
S [32:19]: And then so it's a lot easier to understand.
I [32:21]: It's a lot easier to understand. Is it a lot easier to... um, uh—to communicate physics as well?
S [32:28]: Yeah, uh, I think that's true too, because... uh, like, you can always say like... um... so like, uh... trying to think of a good example... so...
S [32:47]: Like—say, uh... alright, so say you're drawing a free body diagram, and you have uh, several forces here and there. But then, uh, even if you say that uh fnet is zero, you might not understand truly like what the forces involved are. So if you draw a free body diagram and then realize that fnet is zero, then you can use newton's second law somehow. And uh, base it off from there, and keep going on. So even—you might—you might—you might—you might not really understand the question or... if you don't understand the question fully, you can always start from a fundamental principle, and then kind of see what facts you have, and then base it off from there. And it makes it a lot easier to—uh—understand the entire picture, once you start off from like a small fundamental principle. And then go off in that direction.
I [33:43]: Excellent. So, let's uh, return to a moment to discussing ratings and the changes that you've been making in your ratings throughout the semester. At the beginning of the semester you said that you were not getting terribly high calibration scores. Um, was there—what sort of differences were therebetween your ratings and the expert ratings? Was there a trend toward one side of the other?
S [34:11]: Um... uh... oh, wait what do you mean by trend?
I [34:19]: Uh, were you consistently giving lower grades than the experts, or higher grades than the experts, or was it—you were above and below, and there wasn't any uh, rhyme or reason to it.
S [34:31]: Uh, ususally they were kind of above and below. But probably the most inconsistent portion would be... wait, let me double-check. Yes the most inconsistent portion would probably be the content model and content prediction discussion sections. Because now I realize that when they—when they're talking about content models, they're talking about how good your physics concepts relate to the code, and how they're bridged together, so, uh, those two were the ones I had trouble with the most, at the beginning.
I [35:06]: Was it—did you, uh, think they meant something else, or did you just not know what they meant?
S [35:12]: Um, like, I—I was a bit—like because it just said like, "lacks discussion". So I wasn't really sure how much discussion you should have before giving a good score. Because the rubric was a bit vague on what scale you should—how much you should talk about, and how detailed you should be before giving a score of good versus very good versus excellent.
I [35:38]: mm-hmm.
S [35:39]: So I think that all comes down to experience... looking at what types of videos get excellent, as opposed to what types of videos get good ratings.
I [35:48]: What kinds of videos get excellent and good ratings from the experts, you mean?
S [35:55]: So, uh... if you want to get a good rating, I'd say, have a brief discussion of what Newton's second law is. Say, it's force equals mass times acceleration, force equals dp over dt. And then kind of explain the key portion in, um... in the... code, say—I think it's right below—above—right above the position update equation. You want a very good or excellent, you want to go into details, like, of that equation, say, uh here we can also clearly see that constant velocity is present as determined by fnet equals zero, or something like that, that goes into detail about newton's second law.
I [36:39]: And did you get—when you were rating your peers' video, did you get the impression tht people were improving, or that people were beginning to think that... beginning to think like you were thinking on the matter?
S [36:53]: Uh, like, I think perople were definitely improving after Lab 3, I'd say, because uh... uh in the first two labs, a lot of people did not have a... very good introduction, where they stated the problem and then stated the result, and then giving a preview of the major sections. They'd just kind of go directly to some Tracker video, and then base it off—base the code off some other stuff. But... after Lab 3, I saw a lot of videos having, uh, having a powerpoint slide format, having the introduction clearly stated, and then having principles also stated in the other slides. And then based off of there. It was a lot more organized.
I [37:34]: It was a lot more organized. Why do you think it took until lab three, uh, for that to happen?
S [37:40]: Uh... I guess it's just experience. Because, uh, we didn't have a lot of videos to really... guide us through this process. But after lab 3, we probably saw around ten to fifteen videos. Just basing off other peer reviews.
I [37:59]: Excellent. Um... okay, if I have any more... uh, is there anything more about the topic of rating and, particularly, rating over the last semester that you'd like to talk about now?
S [38:13]: Uh... not really.
I [38:15]: Okay, in that case we'll move on to the final portion of the interview. It's—we're running on time. Uh, please open the smaller—the ball rolling example video. Um... I accidentally included another lab four viddeo in that email, please ignore that.
S [38:33]: Okay.
I [38:34]: Yeah.
I [39:04]: Okay. Uh, pause the video please.
S [39:08]: Yep.
I [39:08]: And do you remember this from our first interview?
S [39:10]: Uh, yeah somewhat I do. Well... what am I supposed to do here?
I [39:17]: So please just watch the video, rewind it go back examine any parts you want, and then offer me an explanation of the physics that's going on. And this portion while you're observing it and constructing your explanation, um, I would like you to obey the think-aloud protocol.
S [39:31]: Okay.
I [39:45]: There's a dead period there, and then a tracker video will show up.
S [39:58]: Um... so... um... looking at the... position versus time graph, uh, the ball is rolling in the positive x direction. And um, the—the—curve is very uh linear, which... does suggest a constant velocity. But there is a slight curvature. Which can be... accounted for either... the... point manual point mass, or... uh, some outside forces which can be... which... uh... air resistance is probably nearly negligible, so, in this this case the—it's probably about the frictional force, based on the ball and the ground friction.
S [40:56]: Which may suggest a curvature.
I [41:00]: Okay, can you tell me a little bit more about the role of ground friction in this object's trajectory?
S [41:07]: Uh yeah, so, the force of friction always—is in the opposite direction from the... uh... from the motion of the ball, so... it's the motion—if the momentum of the ball is uh, to the right, then frictional force would be pointing towards the, uh, negative x direction. So—it would definitely slow down the ball a bit.
I [41:34]: And uh, do you think—is that... are there any other factors which might be causing the curvature of the ball's, uh, position versus time plot?
S [41:47]: Um... maybe... uh, on like, I feel like it would be very... negligible, but like, instead of the friction there could be some like, heat transfer. But... I feel like that's um very negligible, like air resistance in this case.
I [42:09]: Why do you feel like air resistance is negligible?

S [42:12]: Uh, because like, the ball is moving at such a slow rate that I feel like, just based on this Tracker position versus time graph, the line's uh, very... straight. So like... the... the... I feel like frictional force would have a much more greater impact.
I [42:36]: Okay. So if the ground fri—so, uh, what sort of changes could we make to this system or its surroundings, uh, if we wanted to eliminate ground friction? Or reduce it.
S [42:50]: Um, so... well like... the frictional force is equal to... the coefficient of friction, and the um, normal force. And the normal force is uh, the normal force is the weight and the gravity multiplied by grav—force of gravity. So, um, one way is to... um, make this coefficient of friction zero. So... we could have like the different—different uh... different um surface, so say like a bowling alley surfave. The bowling alley surface is very smooth and there's a lot less friction than say, like, concrete floor.
I [43:37]: Okay, so, please if you would roll back in the video to the very beginning when the hand is just beginning to push the ball. So if—so imagine that this situation were happening, except it was... um, you know, a very hard polished ball on a bowling ball—on a bowling alley surface. Uh, and if the ball were pushed like that, what would—what would happen in this instance? You can watch it if you like.
S [44:07]: Wait, so uh, if the ball was a—a bowling ball?
I [44:12]: Uh, well, if the ball was the same size—well, forget the ball actually. Suppose we just replaced the concrete floor with a very highly polished, low-friction surface, like a bowling alley. But the same push happened, the hand you know touched the ball in the same way. Would the system be different, and how would it be different?
S [44:35]: Well like, based off the tracker video, um, the [inaudible] was recorded after the hand left the ball for quite—quite some time. So we're just looking at the ball at [this instant?]
I [44:48]: Sure, yeah, in the Tracker portion you're looking at the ball after it's already been set in motion. Uh, but my question has to do with the initial push, when the ball is being set into motion.
S [44:59]: You—oh—so—then you would have to—then you would see some initial velocity.
I [45:04]: You'd see some initial velocity. Would you see... an initial roll?
S [45:10]: An initial roll?
I [45:12]: Yeah. If you pushed the ball on a frictionless surface, would the ball roll, or would it just slide?
S [45:19]: Uh... it would roll because... there would be a... translational kinetic energy and a rotational kinetic energy.
I [45:29]: Can you go into a little more detail about that?
S [45:38]: So like, uh, kinetic energy I believe can be broken down into here in this case, a linear momentum, where the ball is just kind of... say like, in the [experimental?] point-mass system, uh, as seen by tracker, one two three four, moving like that, in a straight line. And that'd just be one half m v squared if you just want to calculate. And there's also rotational kinetic energy, where, um, it's equal to one half moment of inertia times uh... angular velocity squared. And then if you add those two together, you get the total kinetic energy, and uh, some of this can be like, lost due to heat and other factors.
I [46:25]: Okay. And so... how does—now this is a slightly abstract question—so you're talking about energy now. You're talking about kinetic energy being put into the ball. Uh, when in this system is energy added to the ball?
S [46:51]: Um...
S [46:56]: [muttering] energy added to the ball...
S [47:00]: Um, would it be when the hand is giving the ball a push? I believe. [It can make it?] moving from a stationary object.
I [47:11]: Uh, so that's a reasonable explanation. Um, can you go back to the portion of the video when the hand is pushing the ball? I realize it's only a couple of frames, but it might be instructive to see...
S [47:28]: Right here?
I [47:29]: Mm-hmm. And... so before... the hand comes into contact with the ball, uh, can you describe the energy of the ball.
S [47:43]: Um. So... all of the energy is... uh, I'd say, it's uh potential. Potential energy.
I [47:52]: Mm-hmm.
S [47:53]: Like, there's no uh, kinetic energy in the ball.
I [47:58]: And then what about after the hand's come into contact with it? So the next frame... yeah.
S [48:05]: Afterwards there is some kinetic energy, because velocity uh, the ball is now speeding up. Or, not speeding up, but the ball is moving, so... uh, energy is transferred from potential to kinetic.
I [48:21]: Okay. So now I'd like you to go back to the... again, the portion of the video where the hand is in contact with the ball, and we'll do another hypothetical experiment. So, before, I'd asked "imagine if the surface is completely frictionless."
S [48:36]: Uh-huh.
I [48:37]: Uh, and can you—imagine the surface is, is completely frictionless, but this time I'd like you to imagine that that is not a ball, but rather a cube of some sort, about the size of the tennis ball. And if a cube on a frictionless surface were pushed by those hands, like the way those hands are pushing the ball, what would happen in the system?
S [49:02]: Uh, so... because it's a cube, it would just... uh, she's just pushing straight forward, then there would be no rotational kinetic energy. And uh, all the kinetic energy would just be linear.
I [49:18]: Mm-hmm.
S [49:20]: So, the block would just be sliding at a constant velocity, for the entire time.
I [49:26]: And is—is there, is there a way to... make a ball slide on a frictionless surface? And not roll?
S [49:37]: Um... uh, I'm not sure about that.
I [49:43]: Yeah, so when—on the frictionless surface, why do you think... why do you think the cube slides, but the ball rolls? What's the difference between the cube and the ball in that situation?
S [49:59]: Well the... uh... the floor is uh, very horizontal. And the cube's surface is also very horizontal. So I would just assume it slides instead of rolls. Because the cube... the cube, uh, if you just make it a horizontal push, the cube would just... like naturally slide. Like I don't know, it's just... just from like prior mental images, I guess.
S [50:28]: Whereas the ball would've just naturally rolled because there's no... flat surface for it to roll on—for it to slide on.
I [50:36]: Hmm. Um... if you have a moment, uh, if you have a moment—of course you have a moment, we're in the middle of an interview.
S [50:45]: [laughs]
I [50:45]: If you could please go back to your face, uh, we started about five minutes after uh three, so I'm going to take us to five miniutes after four. So we've got about another ten minutes. Uh, do you have a paper and pencil available to you? Or anything to write on.
S [51:06]: Uh, yeah.
I [51:08]: So, I would like you to draw a free body diagram. Um, of a ball sitting on a surface. Uh, receiving a push from a hand directly in line with its center of mass. So in the exact middle of the side of the ball.
I [51:43]: And so for a free body diagram, that would be a, um... you'd be treating the entire ball as a point mass, right, and just summing up the forces on it. Or, you know, identifying all the forces on it. Um... and now I'd like you to repeat that free body diagram for a ball, uh, sitting on a frictional surface, receiving the same push right in the middle.
I [52:24]: And once you've drawn those figures, can you hold them up to the camera for me to see?
S [52:31]: Yes. So—one free body diagram without friction?
I [52:35]: Yes. And—
S [52:35]: And one [on a?] friction surface, and one with friction?
I [52:38]: Correct, and again—and also please include all forces, including the contact force of the floor, and gravity.
S [52:44]: Okay.
I [52:45]: Yeah.
S [53:07]: So what is... like... this... [holds sheet to screen]
I [53:20]: Hold it very still. That's good—yeah. So the top one is... uh, now remember, remember the normal force with the floor. Because if you add those top two on your top diagram, it doesn't appear to me like you have an—an upward force.
S [53:38]: Oh! The—okay, so...
I [53:42]: Yeah.
S [54:25]: This?
I [54:27]: Yes. That's, uh, good. That's what I'd like to see. And now I would like you to do not a free body diagram, but a similar force diagram, this time with the ball considered as a uh, as a solid—as a solid body. As a distinct mass that occupies volume. And please do it with the frictional case and in the non-frictional case. And put the forces on the points on the ball where the forces actually act.
I [54:59]: So instead of a point mass, now it's a solid ball. But it still has forces acting on it at several points.
S [55:05]: Uh, I've never done that before.
I [55:05]: Well, I'd like to see what you do.
S [55:08]: Um... so... if it's—if it's a ball, would [we?] just be the same thing?
I [55:24]: Well... uh... show me what you mean. The same thing.
S [55:31]: Like, the force of gravity would still be pointing downward, and then the normal force would still be pointing upwards.
I [55:39]: But on a solid—on a solid ball, like an actual distinctive ball, if I'm pushing it on the side [gestures] I'm applying a force here. Whereas if it's resting on the ground, the ground applies a force here. Which are in two different parts of the ball. Uh, and so try and give that a diagram and see if that might... um... if this might be an illuminating exercise of some sort.
S [56:34]: Uh, would it still be the same?
I [56:35]: Would there be a what?
S [56:36]: Like, uh, would it still be the same as the previous one?
I [56:45]: Well, the forces would be pointing in the same directions as in the previous ones. Uh, the idea that I'm trying to... explore here is... um... whether things change when you move from a point mass model to a solid body model.
S [57:09]: Yeah, those are like—the point mass doesn't take into account the rotational, [it all like...]
I [57:15]: Yes, the point—the point mass, a point mass doesn't rotate. A point can't really rotate, you know, at least with the physics that we're dealing with here. Um... so in that case I think that's about as far as I want to explore the, uh, this situation right here. Um... I—let me check and see if I took down any notes for followup discussion, and I did not. Uh, unless there's something you'd like to talk about here for our last couple of minutes, um, I think—I think we might call this interview session done.
[logistics etc.]

F.12 Student D, Summer 2017 (Pre)

I [02:15]: Okay. So this is a two-part estimation question. What is the volume of the room that you're sitting in right now? And, if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for the room to fill completely with water?
S [02:35]: Okay, so immediately I'm thinking of something to scale relative to the room that I can sort of multuply by. So I'd go by my own height. I'm about six-two. Um, it takes about two of me to lay across the width of the room. Three of me to lay across the length. So that's about eighteen by twelve feet on floorspace. Then if I stand up... another two of me? So, um, eighteen by twelve by twelve. So I'd multiply that together to get eighteen times twelve times twelve, two thousand five hundred ninety two cubic feet. Um... so that's... um, the answer to the first question, the size.
S [03:25]: And then you asked how long it would take water to fill it? Well, I'd need to know the faucet, the rate of flow, like of how much uh water's coming out per second.
I [03:36]: Mm-hmm.
S [03:37]: So... [is that a thing you could give me?]
I [03:39]: Could you provide, uh, an estimate of the rate of flow out of an ordinary kitchen faucet? Just from your experience?
S [03:48]: An ordinary kitchen faucet?
I [03:48]: Yeah.
S [03:49]: Um... let's see, I guess I think of it in terms of... it takes about... two to three seconds to fill up like a cup of coffee. Like—a coffee cup's worth of uh, worth of water. So I need to be able to convert that to cubic feet... so... every couple seconds... um... how many... I'd need [to get?] my own volume there. So... the volume of a cup of coffee in

relation to a cubic foot, well a coffee cup's about... five inches high... cylinder, so it's... ah man, math's—ahh. Alright I'm going to rethink this.
S [04:42]: Um, one second. How many cubic feet of water are in a gallon? Um... it's like, a tenth of a cubic foot of water? Um... and... let's fill up a gallon of water with a kitchen faucet, I'd estimate that'd take about... ten seconds. So ten seconds times point one cubed feet—so that's twenty-five ninety-two divided by point two—I'd say point one four for a gallon of water, as far as a gallon of water goes. That's eighteen thousand five hundred and fourteen times ten seconds. So that's one eighty-five thousand one hundred forty... um, divided by sixty... divided by sixty again... that's about... fifty one hours to fill an entire room. This entire room.
I [05:56]: Okay. Very good.
S [05:58]: That seems pretty accurate.
I [06:01]: So that was perfect. I was not really interested in the particular answer you gave, just the way that you were doing the think-aloud. And particularly at the beginning, when you were saying, "I am thinking of now", or "I am paying attention to this". That's excellent, that's exactly what we're looking for. So I think we got it.
S [06:16]: Okay.
I [06:17]: Uh, let's go into the section of the interview. Uh, if you'd go to the email I sent you, there is a uh... three, I believe minute long lab video uh, the name of it is some jumble—like an eleven-digit random string.
S [06:37]: Alright. Uh, let me look in the email.
I [06:39]: Yeah.
S [06:40]: And see what you're referring to. Random string? Yeah, six, yeah alright.
I [06:47]: Yeah that's the one. So. Let's set this up, uh what's gonna happen is I'm going to share your screen with me so that get—so that your screen gets recorded. Uh, make sure that you don't have any sensitive information visible when you share your screen, cuz otherwise that's going to go on the recording.
S [07:05]: Will do. Um... uh, let me uh pull up the video and then I'll screenshare with you.
I [07:14]: Yeah.
I [07:16]: And while you're doing this, I want you to act exactly as if you're doing this for a practice or calibration review. Um. Like you did in the previous lab. Uh, so watch the video with an eye to the rubric. I also attached a copy of the expert rubric. And then fill out the rubric, you know—either write down or type down your answers. Um...
S [07:40]: Um, let me pull up the rubric.
I [07:44]: Yeah.
S [08:08]: Alright. So item one... then my notepads [open through this?] [muttering, opens windows]
S [08:27]: Okay.
I [08:27]: Wonderful. So the goal here is to do something like your natural behavior when you're, when you're watching these vid—when you're evaluating the videos. So if you tend to watch it all the way through straight once, then fill it out, do that. If you tend to pause and rewind, go back and forth. do that. Do whatever feels most natural to you.
S [08:44]: Alright. So I'm going to be looking at the organization structure... I like to look at the rubric before, to like look at the things, the checkpoints I'm looking for. So, first I'm looking at... logical structure, organization. An intro that states the problem and the results. Um... good transition. Signposting. Uh, an easy-to-follow video. So that's what I'm going to be grading for on structure.
S [09:16]: Contents... alright. I'm going to be ident—making sure he's identifying the main physics ideas...
S [09:25]: And then prediction, look at how he's initializing the data, where he's getting it from, if that's clear. Also giving discussion of parameters. Alright, yeah, so I've reviewed one of these before. [plays video]
S [09:49]: He's already identifying the models. That's good.
S [10:06]: Alright, so he's identifying where he's getting his data from.
S [10:43]: It's very slowly edited in my opinion. There's a lot of gaps where he's not saying anything. Um... I wish he could have been edited out a little better. Uh, the organization and logical structure seems—he didn't really preview the concept of uh Newton's second law very well, he didn't really explain how it related to the experiment. He just said that we're going to be studying the second law in it.
S [11:15]: Alright.
S [11:44]: Alright so he's explaining where the parameters came from, which is good.
S [12:05]: Alright.
S [12:16]: Again he identifies where the data for his model came from the experimental data in tracker. He definitely could have rehearsed better, as far as delivery goes.
S [12:49]: Alright.
S [13:00]: He explains why... [under breath] constant veloctiy
S [13:17]: He's very in-depth in his explanation of the code, which is good.
S [13:52]: He again identifies what data he used to identify the models. So that is excellent.
S [14:23]: He identifies where the error comes from.
S [14:51]: Alright, so looking at that video as a whole, um, the organization structure I would rate as... um [fair to good?]. He does traverse through the lab logically, it's good to see the initial video, talking about the second law, although very briefly. And then, like, slowly transitions into like showing his video, showing data, showing how he tracked it. Um, however if there is no real signposting, um, and—uh I was never really lost, but he definitely would have benefited from a bit more transitions. So I'd rate organization structure... as... [inaudible]
S [15:48]: Um, content models. So he does identify that we're looking at the equation for Newton's second law, especially going through the code. Uh, he claims... he explains why the system's velocity is constant, which is definitely good. Um... because there's no net force acting on the system, uh, that means no... uh... acceleration occuring. So he shows an understanding of that.
S [16:19]: Um... he doesn't discuss the main physics ideas incredibly in-depth. Like he doesn't outright, uh, state Newton's second law, he kind of skirts around that a bit. Uh, lemme go back... [replays section of video]
S [16:41]: I guess, by highlighting it, it does demonstrate that he knows about the equation. What the law's represented by. Um... he applies these ideas to the problem, like he said, uh, he uh mentioned how other forces were acting on the object in real life and that's why those graphs differed from the uh content—the model that he made in python. Um, and he connects those uh principles and the model and so I'd say that's uh, very good. Um... let's see...
S [17:20]: Content prediction discussion. Uh the data used to initialize the model are clearly identified, he does an excellent job of that throughout the video, he explains where he got his data from, um shows the actual data as well has his graphs. Uh, discusses how the parameters are [inaudible] from the data, he did that as well. Um, and—uh—and his uh discussion of the code, he takes us through the whole loop actually.
S [17:50]: Uh excellent discussion of how the data does or does not fit the model. He does that when he mentions the gap in the graphs, later. Over here [seeks through video]
S [17:58]: Um...
S [18:26]: Alright, what I would have liked is a little bit more as far as the cause of the error. He mentions friction, he also—I want to make sure I'm not cutting him off early [plays video]—alright so what he could have done is also mentioned that there could be some—uh user error, his video, um was just him rolling a jar of peanut butter, he doesn't know that he rolled it at the right velocity, uh the surface could have been a bit more frictionless, yes, there's also the possibility of some air resistance there, but overall he does a very good job when it comes to the prediction discussion. Excellent discussion of whether the computational model does or does not predict the motion of the object observed. That's the one thing, uh, I—he sort of touches on it by talking about how the data matches each other, but not really um—I mean yeah he actually does talk about it when he mentions how—[inaudible] a lot in the code, so overall I'd give him an excellent on content prediction discussion.
S [19:32]: Uh, content overall. Um... I was looking pretty heavily at the structure of the video, and uh the discussion of like the content stuff so... uh, discussion of the uh models. So as far as physics errors, I didn't really see uh any. Um, he did talk about the what if question. Um, cuz I remember the question that was about the x and y axes. And he like, flipped them. He mentions that early on here in the Tracker video. Um, a system—he's identifying his axes. And then as far as uh, let's see... the what does it mean question. I didn't really catch that. Um... I might... be wrong here, uh... [plays video]
S [20:27]: Um, I'm sorry could you give me the what does it mean for this video? What was the question? Cuz he might have answered it but not identified it.
I [20:34]: Uh, yes, that is a good question. Give me just one moment let me bring it up. Don't have it off the top of my head. Uh the what does it mean question for this lab one—view lab one...
I [20:55]: Uh, the what does it mean. Your observations can be predicted well by your computational model when you set the net force to zero in your model. Is it possible for you to say how many pushes and pulls are added together to give zero net force in the case you observe? And discuss.
S [21:13]: Ah. I remember this one. He... so he mentions some of the forces, um... that could occur. He didn't do a very good—he didn't really actually like directly address the what does it mean question. He didn't really address the ability to count them up. So that... was completely missed. Uh... if I'm not... mistaken. So while he does the uh first two pretty well, the lack of a what does it mean um, I put that—I guess as a... it's one out of the three criteria. So... out of—I guess I'd put that as a good instead of a very good or an excellent. Because he does a good job with the other two. Um, so... good—um—production delivery.
S [22:03]: Um so the lighting and resolution's alright, and the audio quality is okay—the audio quality is good too, there's no [shaking?] video. But overall the production quality isn't great. Um, overall the audio quality is—uh the production quality isn't that great becasue there's a lot fo pausing, it wasn't entirely polished, I feel like he could have done a better job of rehearsing and like cutting his delivery a little bit more, as opposed to leaving—it feels like uh, you know, pausing and a dead space in there. Um, but other than that, everything else about it was, uh, well-done, uh, and again there's also really a lack of transitions which kind of [inaudible] stuff up.
S [22:46]: Um, but I'd give it a very good on production delivery because he's really only missing um, a—like a few transitions, and uh the delivery was very clear and easy to follow. So, uh, I'd give it a very good. Like, grading of the uh, video.
I [23:06]: Okay, very cool. Uh, are those the final grades that you would like to... assign?
S [23:11]: Um... [sighs] let's see. Look at the rubric, uh, back through. I might... just a sec—content prediction is excellent. [muttering, reading] ...good... ...very good... content models... I'd say, I'd say his organization structure is actually fair. I wouldn't actually put good on that. But other than that that is my final. Fair, good—fair, very good, excellent, good.
I [23:45]: Can you tell me why you revised the organization structure down?
S [23:49]: Thinking back on it, I dunno. Um... I... realized that like I was—when I was opening my analysis of the video, I didn't want to grade too harshly. But I realized that uh, organization is a very com—compartmentalized, like, aspect of the video—like of the video. Like... grading. Like it's [hugely?] just about how it's uh, presented to you. How easy it is to follow. And while it did have, like, uh an easy flow, like I was never lost, it had no signposting at all. No uh transitions. And uh, he doesn't really give a big preview in his introduction either. Statement of result is... um... but everything else is like kind of uh present, so I didn't wanna like give him a poor grade. But it wasn't great either.
I [24:38]: Okay, so you've touched on a couple interesting themes that I'd like to come to in just a moment, but for now please keep your screen up. I'm emailing you the expert evaluation for that same video.
S [24:48]: Yeah.
I [24:48]: So I'd like you to read that, review it, and then tell me about your thoughts.
S [25:07]: Fair. Alright so they gave the same grade as me... on organization. We differed on... uh, let me pull up the email [inaudible]. We differed on the, uh... content models, so they said there was no explanation of the physics principles... in the system. They gave him a poor. Uh... I don't know, he applies the ideas to the problem and identifies like, the like the relevant equations and uh... models, so I don't—I—fair I'd understand, I don't know about poor, but I do think I probably overrated him on that.
S [25:50]: Um... where they say good, I said excellent on content prediction. So they gave him a good video. [but they didn't say?] very good observation, good account of friction. Account of initial condit—was buried in a disorganized explanation of the python code. Um. While it's fair for them to say that maybe that was muddled, I would disagree with how they graded that actually. Just because the delivery may have been disorganized, that's something to criticise more on the organization or production delivery section. I don't think that should bleed into the content prediction section at all. Uh, when they themselves say very good, I think they should at least give a very good on, uh, the content prediction.
S [26:32]: Or, as I said it was excellent. Um...

S [26:35]: The uh... we gave the same grade on um, content overall. Inadequate WDIM explanation, no motion net force connection. Um, I wouldn't say there—he didn't say—I'd disagree that there's no motion and net force connection. He explicitly states that because there's no force, the object—the object's velocity doesn't change, so, I mean... uh, he does mention that. Um, but good WI, physics principles appear to be soundly understood but not clearly expressed. I definitely agree with that analysis. He clearly understands the concepts but he might have not communicated them as explicitly as they may have like.
S [27:17]: Um, and then production delivery, very good, we again agree. Except for hard-to-read walls of code. I mean there really isn't a better way to present the code in my opinion, because it's not that big—not that big of a flaw, it doesn't like... it's a pretty simple code, so it's pretty easy to follow and he talks through it, so I disagree that it's just hard to follow. But, um... but we agreed on the grading, so...
I [27:43]: Okay, very cool. Could you please, um, switch around to your—so I can see your face.
S [27:50]: Sure. Um, lemme turn off screen sharing, it's—alright.
I [27:57]: Okay, very cool. So now we're going to have a more traditional back and forth sort of interview about what just happened and the decisions that you made and the ratings that you made. Um, so, firstly, uh... how does this video compare to the other videos that you've been watching?
S [28:16]: Um, it's around average compared to the other videos I've been watching. Like it's not—not to toot my own horn, but we'd be asked to relate them to ours a lot. I'd say my video is more well done than his, but not by a huge margin. Um. But mine—it's definitely like worse than mine, but like better—a few of the one's I've seen are a lot vaguer, and like didn't discuss physics at all. Like they just kind of took them through their lab data. So those were the worst ones I say, and that was about—that one was pretty decent, I saw a few that were better and [inaudible] that I also thought were better.
I [28:53]: Um, and you also mentioned earlier that you didn't want to grade too harshly. You mentioned a couple of times that you didn't want to give them a poor grade or grade too harshly. Could you talk a little bit about that.
S [29:06]: In my opinion, um, if someone meets the requirements, they deserve the credit. I don't—I don't want to come off as expecting someone to have to like, exceed any expectations I have for them to earn the points that they deserve on a rubric. So I lean towards uh, as far as not wanting to grade harshly, I lean towards giving someone credit uh for something, rather than—if it's borderline, rather than um... not giving them credit. Uh—
I [29:38]: And do you think—how does that attitude do you think compare with how the experts tend to approach grading?
S [29:45]: I think the experts are harsher.
I [29:47]: Could you talk a little bit abou that?
S [29:50]: I feel that... like... even when, even when their descriptions seem to—describe, describe a lot of the good things that the person did, and that they succeeded in doing, they don't match the grade as accurately—I don't know, I don't really understand why they dock as many points as they do in some of the cases. Like, uh lemme pull up the other grade again.
S [30:14]: Like again, um, they said uh, it was only um... good when they themselves go on to describe it as very good in certain aspects. And uh, I feel if they're good enough to reach very good in their eyes, uh, in broad description, then the smaller type of grade should also be very good.
I [30:35]: So there's sometimes a mismatch bewteen what it appears the experts are saying in the comments, and the ratings that they're actually giving.
S [30:42]: Yes. I would absolutely agree with that statement. And I also feel that, um, the experts lean towards not giving them points. Like, they—I feel like they, they require a lot more... their criteria um, doesn't... impart upon people how explicit they seem they want them to be about certain things. So, um, whereas if I see them describing the physics concepts and applying to the model while they're talking instead of explic—explicitly stating "this is how this model matches newton's second law, because this part of this equation is this, uh, part of the vector of the object's motion" or whatever.
I [31:23]: Mm-hmm.
S [31:23]: Um, I feel like that might be what the uh experts are looking for, but they're not really imparting upon—imparting that upon the student in like the experiment. Um, I think that credit should be given when a student, like, shows that they know the concept and that it's like apparent.
I [31:41]: Hmm. So why—why do you think the experts, uh, tend to do that?
S [31:45]: I—I—I really don't know, I mean their motives, I guess. They just have harsher criteria personally than me. Um... their, their definition of fair and good differ from mine. The reason that I... uh, for like—this because, like the—um, the scale is from uh poor to fair to good to very good to excellent. And so, I'm not really looking at it in terms of those words, I'm looking at it more in terms of, um, fair is the worst and excellent is—excellent is the best. And I just kind of think of it as a grading scale in that [line?]. I'm looking at how far left or right my grade is, really, as opposed to, "Is it—is it fair, or is it very very good?" Like uh, if what I'm saying makes sense to you. Um, I'm looking at the grading as more of a s—like a number line. And, in this case, from poor being uh, like an F, and Excellent being an A.
I [32:48]: Right. And so you think... and how, how do you... so you have, you have this number line so there's some scale, right, and you'll judge the quality of some element of the video, and it belongs somewhere on the scale. Um, do you think—how did you decide to set you know, where excellent is, and where poor is? How bad do you have to be to get poor, how good do you have to be to get excellent?
S [33:16]: I... sort of looked at like, how much of the components of uh, like the rubric were in the boxes on the right and left. How much it matched. And I just said like, oh if it matches close—more closely the advanced versus the marginal, then I'm going to put it closer to excellent. And if it matches, like, most to all of it, then I'm going to put it in the excellent column, rather than the very good column.
I [33:41]: Uh, do you find that your ratings are... um, frequently... much different than the expert ratings? Not just in this video, but in the whole set of them that you've graded?
S [33:51]: Uh, in this one honestly mine were pretty close. I think I did a pretty accurate job. Um... minus the two or three differences we had, where I'm off by one or two. Um, but in the... videos when we reviewed in Webassign, I think it was, that we did it—
I [34:10]: Mm-hmm.
S [34:11]: —for the first lab, I'd say... they weren't as close as they are right now. They were—mine were a little bit more high than the experts were.
I [34:21]: Ah, yeah, I'd like to talk about that a bit. Is that—how does that relate to not wanting to grade too harshly, do you think?
S [34:30]: What, the expert opinions grades?
I [34:33]: The fact that for your lab one evaluations, your ratings tended to be higher than the expert ratings.
S [34:39]: Um... so how did that factor in for me not wanting to grade too harshly? Well I would—since doing that, I do—I started grading more, more uh leniently. And that made me swing more to the slightly harsher. But in this, I think I'm kind of an average between the two. Um, where it's—um—like I'm definitely, they definitely have to earn their points, but I'm trying to see where they're succeeding, sort of. As far as not wanting to grade harshly.
I [35:11]: And do you think your peers, the other students in the class, uh, have the same sort of attitude as you do about that?
S [35:17]: Um, I dunno. I mean, per—I'd assume so. Like I assume they're all students who... sympathize with fellow students in some shape or form. But...
I [35:29]: Sympathizing is an interesting phrase. Can you talk about that a bit.
S [35:36]: Um, one, um... [muted conversation with someone offscreen]
S [35:41]: Um... so... sympathize. I dunno. We all go to Tech, Tech is hard, so... people don't wanna make things like, people know—realize that if they're struggling everyone else is probably struggling too. Might as well, like, not make it harder for people than it needs to be if they're earning their grades.
I [36:01]: Hmm. So that sounds like a fairly, um, cooperative attitude, not a terribly competitive attitude.
S [36:08]: Yeah. I mean that's just for me personally, though. I don't really know how other people are responding to this.
I [36:16]: Very good. Alright, so, uh I will respect your time. Let's—this has been a very interesting section of the interview, uh, let's move on to the last one. So if you go to your email again, there's a ball rolling example dot mov. Um, and, I'd like you to share your screen, watch the video, explain and then come up with an explanation for the physics that's going on, and the findings in Tracker as if you were preparing an explanation for a lab video.
S [36:46]: Alright.
I [36:47]: And then we'll have a discussion about that.
S [36:50]: So, you want me to explain it like I would if I was preparing for a lab video.
I [36:54]: Yeah. Watch the video and then do a think-aloud where you're pointing out to me what you notice and what you're thinking about. And then offer to me an explanation of the physics undergoing it, and then we'll have a discussion.
S [37:08]: Alright. Let me share my screen with you, and we can do that. Share screen...
S [37:16]: Alright. So in the video you sent me... tennis ball, uh obviously right is going to be the positive motion. There's an initial force, but after that no net force acting on the object. Relatively frictionless floor. The object seemed to move at a pretty constant velocity, but let's [roll back the tape?].
S [37:42]: Um... yes, the object, uh the ball seems to be rolling at a relatively constant rate.
I [37:50]: And if you'll continue the video there's a tracker analysis at the end, as well.
S [37:55]: Continue the video track—oh, okay, let me go back to it.
S [38:04]: Oh. Let's see. Oh, perfect. Okay so the ball is slightly slowing down at the end, seen by the slight bowing of the graph from... we're curving slightly. But it's definitely relatively constant, almost a constant slope. Um... I'm seeing a good job hitting the center of the ball each time. Um. Again the the y—the y-axis is irrelevant, the x-axis the right seems to be positive in his—this scenario. Um...
S [38:38]: Um... so... other than that, yeah. I mean... constant motion. Slope of the graph is relatively constant because there's no real acceleration because there's no net force acting on the ball.
I [38:52]: So we'll go right into the discussion. Keep your screen up because we might be referring back to the ball rolling video. Because I would like to talk about um, so you made a phrase... uh, "a relatively frictionless floor". Could you explain why... why you mean that? And what you mean by that?
S [39:11]: So, if you're trying to analyze a ball that's moving at a constant velocity, you don't want other forces acting on it, cuz that cause—cuz a change in force causes a change in acceleration, meaning its velocity changes. So if...
I [39:26]: So—
S [39:29]: And. Hmm?
I [39:29]: Mm-hmm.
I [39:28]: Sorry, continue. I keep—there's a lag, so I keep interrupting you.
S [39:32]: Alright. So, if... we want to know changes in acceleration, we want to minimize the amount of forces that are acting externally on the system. One of the main ones that we see in like labs is we're assuming that it's a frictionless, uh, surface with no drag or air resistance, right. Most physics problems—like a lot of physics problems you're allowed to assume that. If you're trying to look at like look at a system like this where it's constant velocity. So, um, the floor is flat, seemed pretty smooth, it did look like a bit dirty so it's not completely frictionless obviously, uh enough so that in this reference frame of the video, it's not a—friction doesn't seem to be playing a large role in the object's movement.
I [40:14]: Right, and why are you justified in saying that friction is not playing a large role? I'd like to connect this to the um, plot in tracker.
S [40:21]: Okay. So, if—the ball—it does play a small role, as seen like I said, it does curve a bit towards the end. Slows down a bit. But the slope is decently constant. And [take also into account?] that... uh, his clicks might not be perfect, uh, on the side of the ball anyways.
I [40:41]: Yeah, so if his clicks were not perfect, might that be a contributing factor to the downward curve of the graph?
S [40:48]: Exactly. Yeah if he accidentally clicked a little too far to the left, the ball wouldn't look like it moved as much in the graph as it did in the actual motion.
I [41:00]: Yes, I'd like to explore that concept a bit. So imagine that he did, you know, um, imagine that you were a little inaccurate every time you clicked to find the center of the ball. What sort of plot would that produce?
S [41:17]: What sort of plot?
I [41:18]: Yeah. What would be the shape of the line, the traj—the position versus time line.
S [41:23]: Yeah it'd be more jagged, as seen in these small fluctuations he has, so it's not going to be as accurate.

I [41:28]: Okay, very good. And so, uh... looking at the ball right there at that point in the trajectory, uh, can you list for me the forces uh which you think are acting on the ball?
S [41:42]: Uh, force of—uh the weight force, so mass times gravity going down. There's the normal force of the floor pushing up on the ball going upwards. Um, it's... currently there's no force in the x—it's just remaining in motion because it was already placed into motion by uh the force applied, which happened earlier, but it's not currently acting on the ball.
I [42:06]: Mm-hmm.
S [42:08]: So really only forces that cancel out in the y are occuring. That's is why there's a constant uh zero acceleration. Near zero acceleration.
I [42:15]: Very good. And so I'd like to return a little bit to the discussion about, um, the... friction between the ball and the floor. So—
S [42:23]: Oh! Oh! Friction was acting on it—I was assuming friction was zero. So pushing back on the ball to the left, in the negative x would be friction. But, uh—
I [42:32]: Right, yeah.
S [42:32]: —we're assuming that—we're assuming it's negligible.
I [42:36]: We're assuming it's negligible. So... we're assuming that friction is negligible here, uh which is—
S [42:45]: Well I'm assuming this is what—like, like a lab video that's in preparation for creating like a model in python.
I [42:50]: Mm-hmm.
S [42:51]: In Vpython. Where we didn't really consider friction.
I [42:54]: Right, so—
S [42:54]: But, um—
I [42:54]: Yeah.
S [42:54]: If you're just looking at the analysis of the ball's motion, purely, then yes friction is playing a role.
I [43:01]: Yes, that is—yes, that is what I'd like to get at here, with a little bit of a further discussion about the friction. So you say the floor is mostly smooth, but a little bit dirty, and so it's clearly not perfectly frictionless.
S [43:16]: Yeah.
I [43:16]: If you clean the floor, or waxed, or you know—covered it with glass or something, would that produce a different outcome here? Would that change the role of friction?
S [43:28]: Uh, friction would be even further minimized, yes.
I [43:31]: And what sort of effect would that have on the position vs. time plot?
S [43:35]: Uh, position vs. time plot? It wouldn't curve downwards at—uh, as much. Like... where it's slightly bowing out, you'd see a flatter, uh, slope of it.
I [43:46]: Very good. So would this analysis apply to, uh, something which is sliding across the floor? Rather than rolling? Is there—is there a difference between those situations? Particularly with the role of friction.
S [44:03]: Uh, I feel like something sliding along the floor—friction probably plays a larger role. Because there's more of the object... in contact with the floor. Like, at all times, if you're just dragging something along the floor, um, friction's going to be pushing back harder than it would if it were just rolling along.
I [44:25]: Okay, and so is that generally true—if something rolls, the friction is different than if it slides?
S [44:32]: Uh, I would say so. Yes. But I don't—I'm not an expert [laughs] I don't know for sure.
I [44:38]: Uh, well, I mean—if you have a thought on that topic, I'd like to explore it.
S [44:43]: Yeah, I mean, um... like—thinking about, like, if there's more surface... for the uh, for it to be like acting on the object, so if it's like, just like a plate that you're sliding around—along the floor, like, just like a block, um, then I feel like there's more contact between the object and the floor and therefore more friction, but I don't know how accurate that is. Um, and then on top of that, if something's rolling... like... it's... moving more easily. There's only like one point of the object that's touching at any certain instant in time. So there's less friction [inaudible].
I [45:25]: Very cool. Well, that is the end of our hour. I think that was an extremely productive interview.
[logistics etc.]

F.13 Student D, Summer 2017 (Post)

[logistics, etc.]
I [04:35]: I'll tell you about what we're going to do during this interview, and then just go ahead and get started. It's going to be essentially the same as the interview we had at the beginning of the semester. You had what I hope was a rich and fulfilling summer in the interim period. Um, and we're gonna get you to watch the same video, do the same evaluation, and discuss the same physics topics so we can explore what sort of changes have happened in, you know, your own mind over the course of the summer. Uh, we're not going to start out with the practice problem because we're pretty sure that you have a solid grasp on the think-aloud protocol, so we're just going to get right into it. If you could please watch the lab video, uh, that I emailed you, uh, please share your screen while you're doing it. Uh, and... watch and evaluate and assign a rating and a comment to each item on the rubric for that video as you normally would.
S [05:25]: Alright.
I [05:26]: ...while participating in, um, the think-aloud protocol.
S [05:30]: Alright.
I [05:31]: Very good.
S [05:32]: Let me pull up the... rubric. And then I will pull up the, uh... [operating computer]
S [05:49]: Alright. Rubric. And... video... let's, alright. And so I'll [participate?]
S [06:06]: Alright so immediately what I'm noticing is a... visually, he doesn't have like, presentation setup like the title, and, um, lab, most introductions over the year that I looked at suffered a lot from disorganization. They did it because they didn't have like a clear organizing structure, this dude's just showing straight from their tracker data.
S [06:46]: Alright. He said our axes are straight, but he didn't say right is positive, left is negative. Which you should state when you're uh, finding your reference frame for... physics.
S [07:00]: He does describe, uh, the units that are important. Um, uh, let's—uh—like—the objects being recorded. He noted the diameter. Alright, perfect, there he goes mentioning positive x.
S [07:25]: This, uh, part that he's referring to right now if I remember is part of the what does it mean, what if part. So what if you rotated the axes. He's getting to it a little bit early, and not explaining the actual question he's trying to answer there.
S [07:46]: That's valid.
S [07:55]: He seems to do a good job of keeping the center of the circle on the center of mass of the peanut butter.
S [08:06]: There is a little bit of foreshadowing there that he just—when he says he'll discuss why the slope is relatively constant. Well, he didn't say slope. Hopefully he'll discuss the slope.
S [08:20]: Again, um, the... organization structure and the way it's delivered is a bit poor in that it doesn't... have any signposting. It doesn't have, like, a title card with code with the relevant parts. He's kind of just clicking around on his desktop, selecting parts of the experiment.
S [08:47]: He uses SI units. He doesn't say where he got the mass of peanut butter.
S [09:04]: Um, this shows a little bit of inaccuracy in his tracker setup. Um, the position... would be zero zero zero, um, if he had the axis, uh—if he had the origin directly on the starting point. That's an understandable margin of error. Negative point zero zero three meters isn't much.
S [09:38]: It'd be a little helpful if he referred to, um, instead of just the x column and t column, what the actual values he's using are. So, position in the x-axis, and time in seconds.
S [09:54]: Alright, so... again, he denotes where he got some of his data from, which is helpful. Which helps his content models.
S [10:26]: This shows a lack of rehearsal on the delivery side, as far as narration goes. He stumbles over himself a bit. He could have just cut the video and started the recording from the point where he stumbled. He also didn't really discuss correctly the forces. Uh... he should've just stated that at the start, the force is zero because an objects at rest tends to stay at rest, and there's no outside forces acting in an imbalanced net force on it. Um, he just said they're invisible forces we can't see, so it'd be helpful for him to more specifically denote where the forces are coming from. Um, and also if he had a more full explanation than that, he'd have a more fluid narration.
S [11:14]: That's good, he's identifying the important model. Um, one thing I will note is that at the start of the video, its helpful in the introduction you're supposed to preview the major sections and, like, the major... uh... mod—major models that you're going to be using, and major physics principles. And he's just now getting to—referring to Newton's second law, uh, more than halfway through the video.
S [11:47]: This is also, uh, helpful.
S [12:06]: So he should explain that delta t is a small timestep... uh, and explain that the iter—they're iteratively calculating the uh position of the uh peanut butter over time by using, uh... the derivation of Newton's second law, um, in showing that position plus uh position times a small timestep over time will give you uh, a valid, uh, position update.
S [12:32]: Um, he also should've done a similar thing with systems. With the system's velocity. But moreso focusing on the fact that it is constant because, um, there's just one initial force acting on it, and—there's not much in the way of drag or friction because it's a relatively frictional [sic] surface uh, under the peanut butter that would be helpful. As far as actually discussing the phyics concepts behind why the peanut butter's moving the way it is.
S [13:16]: Um, he should discuss more why it's straight.
S [13:29]: So again he's showing where his data comes from, which is helpful. This is a decent graph of the position over time, and comparison between them.
S [13:44]: They're relatively close. His model's decently accurate, but there's a pretty big split towards the end.
S [14:00]: Okay, he has that correct.
S [14:12]: That is a sound discussion of the physics behind the movement. Um... he didn't really have to mention drag, because it's just a... a very—the—the—it plays a minute, pretty much negligible role. Um, he... could have explained how he could have made it better. Um, we'll see if he does that later. He could.
S [14:38]: Okay, so there's no conclusion, really, he just kinds of ends explaining the graphs. Um, there should have been a general wrap-up, addressing of the what if and what does it mean questions. Um, as well as a more in-depth explanation of, "here's what I could have done differently to minimize friction", maybe make the graphs more accurate. He didn't really explain... uh, he didn't really use any, um, mathematical demonstration of how accurate his model was. He could have presented either the r-squared value, I know excel does that automatically, um, if you check a box. And as shown—uh, if it has a point nine something, it's a pretty high correlation. That would also have been helpful. But that's uh, my think-aloud on the, uh, video.
I [15:24]: Alright. Now, uh, could you please uh look at the rubric that I emailed you and actually assign a rating to each item on the rubric.
S [15:32]: Alright. Uh, I can do that right now. Um... can—do I just do that out loud.
I [15:37]: Uh, I would prefer if you typed it up, maybe in a notepad document or something, so that you could email it to me.
S [15:48]: Alright. Lemme get notepad up. Alright I can do screensharing...
I [15:51]: Yes, thank you.
S [15:58]: [muttering] rubric, here's notepad...
S [16:00]: So organization/structure, um. So he lacked an intro—well he had a very, very short intro, didn't really give a—lemme go back. Am I allowed to refer back to the video?
I [16:17]: Uh, yes. Please. So what I want you to do is to treat this as if you were doing an actual peer evaluaiton. So watch the video, go back and forth, type up whatever you like, in whatever way is most natural to you.
S [16:35]: Okay, so he does say the purpose, but it's not really in-depth. He just says the purpose of this lab is to use Newton's second law to observe a system in constant motion—
I [16:47]: And I'm very sorry to interrupt but I've lost your video feed.
S [16:51]: Okay, um...
I [16:53]: I'm not sure what you did last time that fixed it.

S [16:57]: Let's try this again. Awesome, is that working now?
I [17:01]: Yes, now it is. Yes.
S [17:03]: Uh, so he did have one of the more important parts of the intro, in that he gives like the purpose. He gives no preview of the major sections, though. Um, no... real... uh, overview of the actual equations for Newton's second law, or like a general explanation of the actual principle, just the fact that we're... exploring the principle. Um, and there's absolutely no conclusion and very few, zero—there are zero transitions and signposting. I'm never lost in the video, but overall that the organization structure is poor.
S [17:42]: Uh, for the content models, um, he does identify, um... the fact that Newton's second law is important, and he does use the correct equations for updating velocity and position. Um, he does lack an in-depth discussion of the main physics ideas, he just sort of briefly touches on some of them. Um, he does sort—he does describe the application of them in the code, though, so I would give him a fair on this section.
S [18:16]: Uh, the content prediction discussion. Um, so the data—he does do a very good job showing where the data he's using is coming from, minus the way—he doesn't describe the mass of the peanut butter jar. So he does not define where that comes from. He does however identify where all of his tracker data comes from and uh, where all his graphing data comes from. Um, he doesn't really discuss the parameter adjustment that much, he sort of talks about how, uh, fnet is left to zero but his discussion isn't exactly... uh, in-depth there. He just kinds of mentions it.
S [18:51]: Um, he does talk about the fit of the data relative to the model, but not mathematically, and as far as discussion of whether the computational model does or does not predict the motion actually observed, he does—he—he did a decent job talking, talking about that. I'm giving him a fair in the content prediction discussion.
S [19:11]: Um, overall, there aren't many errors in his physics, um, as far as like the code and actually, uh, functionally, uh, but his discussions of the physics concepts aren't... super informed. He doesn't have uh, a strong grasp of why—why newton's second law is important, and, um, how to actually describe to the viewer beforehand. Uh, the what does it mean question he never even mentions, I don't believe, I'll double check that though. Um, do you—do you know what the what does it mean question was for this video.
I [19:49]: Uh, the what does it mean questions is—uh—is uh, how can we... uh, can we identify all the individual forces on this—uh, what can we say about the individual forces on this object, given what we've collected from the motion data?
S [20:10]: Okay, so he does talk about that, he doesn't—clearly identify where he's talking about that, but he does mention that there is friction acting on the jar, um, and he does mention the applied force of the hand. Um... he doesn't really talk about if we can... calculate all of them. Um, at least—let's—[to our knowledge?] [plays video]
S [20:35]: So yeah, here's him talking about friction.
S [20:47]: So there here's talking about the fit again. So, overall, his content um, I give, again a fair. It's not great, but uh, there's nothing glaringly wrong either. Um, and then production/delivery, he had decent resolution, his audio was clear, uh, his narration could stand some improvement, and the visuals aren't that, uh, aren't that well-thought-out. Um, I would say I'd give his production delivery in general, he hits... more than half the checkpoints, so there's nothing distracting, so I'd say his production is good. Definitely not very good or excellent. Um... content overall... fair... and, uh, usually I go back over and just double check.
S [21:34]: So organization structure, um, yeah he doesn't state the result in the intro. Um... doesn't preview major sections, doesn't transition or signpost, I'm never lost but that's not enough to account for missing every other—alright, so every other section on there, he didn't do well. So poor.
S [21:54]: Um, so content models, he does identify some of the relevant models. Mediocre discussion of the ideas relative to the problem. And physics, um, he doesn't actually really, yeah, he doesn't do a very good job of connecting it to, um, but... overall, fair.
S [22:16]: Um... I'm gonna drop the production delivery to fair, because actually thinking back, the narration was pretty stilted. Um, ehh, no, it's good. Uh, it's good.
S [22:30]: So those are my ratings for the video.
I [22:34]: Okay, excellent, if—when you have a moment, maybe after the interview, if you could please email that document to me.
S [22:41]: Alright.
I [22:42]: For now, I'm emailing you the, uh, instructor rating for that same video. Should be in you—in your inbox momentarily. And, I'd like you to look over that and react to it as if you were looking at the expoert evaluation of a practice or calibration video.
S [23:02]: So this is, actually, very surprising to me. Um... usually my ratings are too high, compared to uh, the ones given. Um, so... what [are they?]
S [23:15]: So fair, no results in intro—so they said his description of lab's purpose was good, um, wi—wdim, what's wdim? Oh, what does it mean and what—yeah. Yeah so they said it should have been signposted. He didn't really signpost it, which I said... um, fair for organization structure. Uh, yeah they gave him more credit than I expected him to. So we flipped on fair and poor for content models, and organization structure. Uh, they said no explanation of physics principles and the connection between them, I guess I gave him a little too much credit... um, for more of the mentioning, as far as—as opposed to the actual explanaiton of the physics principles. Um, content... model discussion, they said he had a very good observational/model comparison, I... I would disagree. Uh, I thought it was very... basic, and like, um, not—not very far in-depth. Uh, he... does do a good job of accounting for the role of friction at the time it really kicks in.
S [24:24]: Um, and the initial conditions... the account of the initial conditions calculation was buried in a disorganized explanation of the python code. Uh, I would agree with that. Um, which is why I'm a little confused as to their good rating. Um, they sound more critical, uh, of the rest...
S [24:43]: Uh. Subjects—um, so, content overall, good, inadequate what does it mean explanation, again I agree, no motion net force connection, which is what I discussed earlier when I was watching it. Good... what if, um... yeah, he did, he was correct in that, um, on the x and y, said that that was sound. Um physics principles appear to be soundly understood but not always very clearly expressed. I think that's a... decent explanation of uh, of... the video's overall vibe. Um... I guess I would—I guess I swung a little hard on the pendulum the other direction after my reaction to the last grading, where I was way lower than them on each of them. Uh, I've—I was close, I feel like, on these those. I was only one away on most of them. The one thing I definitely disagree with is production quality. Um, because uh I don' t know—one—as I took the class, the production—I guess my... um, expectations for the production quality of the videos went up because uh, most people switched to using uh, power—well-organized powerpoints with narration backing it. Um, and I found those a lot easier to follow, and they did a better job of like presenting the models and topics. Um, and text that you can read along as well as listen, uh this video had no real visual enhancement. Um... disorganized, and I thought the narration was a bit stilted. But... uh, I would've—I would put it as good, not very good, but other than that, yeah, those are my reactions to the... ratings.
I [26:27]: Okay, excellent. Uh, so you've raised some interesting points that I'd like to discuss with you for a little while. Uh, could you switch the camera back around?
S [26:36]: Sure can.
I [26:37]: And... once again, I've lost [your video feed]
S [26:41]: [there you go]
I [26:41]: Okay, excellent. So you mentioned um... that you were surprised. Usually your ratings were higher than those of the experts. Uh, and you mentioned something about reac—swinging on a pendulum, reacting to your last grading. So, can you talk about that for a bit?
S [26:58]: Um, so, uh... in the first interview we did, and throughout the course, um, my uh grading whenever I looked at my grading relative to the uh, instructor grading, I always found them a bit unfair and super strict.
I [27:18]: Mm-hmm.
S [27:18]: Uh. Where I discussed in the last—my last—um, interview, I thought that in the description they said stuff like very good, and he did a good job on this, and still giving him a low rating.
I [27:34]: Mm-hmm.
S [27:35]: In this situation, I think it's—what they—they give better ratings relative to their description, I guess I'm putting a little bit too much on the description itself. Um, but for me, I think... uh, after like seeing how I got poor grades on a lot of the, uh... grading assignments, um, on the labs, I saw that like they were looking for a lot more stringent um, requirements for uh, higher grades on each of the sections. So I think that's just, that's how I reacted. I was a little bit more stringent.
I [28:10]: And so you found as you got more stringent as the semester went on, um, how did you... how is that reflected in the way that you thought about videos? What did it mean to become more stringent?
S [28:25]: So, basically, um... at, at first when I was first grading them, I looked at the rubric and I, I was... looking for reasons to give them credit, basically. Uh, if I saw what they did, uh, fit in any way, one of the descriptions on the good side, I'd give that as a plus to them. Like, that's good, they're good on that. Um... let's keep going.
S [28:52]: Um, whereas, as I went on in the course, my grades more required repeated, and like consistent achieving of the criteria. So, um, like... for example [operates computer] when it says uh, like, does a good job of connecting the fundamental um, physics concepts and the concepts in the model, um, instead of then just saying, oh and the model fit pretty well, it did a good job of predicting the movement using Newton's second law, I'd require them—like—to get credit, I'd require it to be more, "Newton's Second Law states this, um, this is why it connects well with this model. Like, the object's velocity here in the python graph is in the equation here, and et cetera." So, like I dunno, just more consistency was required.
I [29:47]: And why did you think uh, you began the semester with that attitude of trying to find reasons to give people credit? Where did that attitude come from?
S [29:57]: Uh, just, mainly sympathy towards my fellow student. Like I thought that if, uh, if I were in the situation of the person being graded, um, I'd want the grader to... give me credit, uh, whenever possible—whatever [is possibly due?]. Um... and I was just trying to be a reasonably fair, um, but not like ridiculously easy grader.
I [30:20]: And how, how would you describe your feelings of sympathy now, at the end of the semester? Or rather, you know, back around lab four, around the end of the semester.
S [30:31]: Um, so... I still find myself uh... once I found out how the grading was figured out, where... you were—your decisions are weighted based on how well you matched the instructors' guidelines...
I [30:51]: Mm-hmm.
S [30:52]: ...um, in the calibration activity, um, I mean I still had like—internally, I still fell like, oh, like, these are my fellow students, but I more thought of, how I'm going to like most accurately match the system, as opposed to how I'm supposed to grade these.
I [31:10]: And... um... tell me a little more about why you think the instructors—so, you're interested in matching the system. You're interested in anticipating what the instructor responses are going to be. Um, what do you think motivates the instructors to set the responses they did?
S [31:30]: Um, I don't... so... for me it's kind of hard—it's hard for me to really know, considering... I dunno, I don't really have much grading experience.
I [31:39]: Mm-hmm.
S [31:40]: Um, I guess that...the biggest um... ideal that instructors are trying to achieve is objectivity.
I [31:52]: Mm-hmm.
S [31:53]: So.
I [31:53]: [And so do you feel]
S [31:54]: [For me I think my problem was] there was a lot of gray areas, to me, and so I was more lenient at the start. And for them, it's... is it right, or is it wrong.
I [32:05]: So this last, um... instructors see less of a gray area, is that what you're...?
S [32:12]: Yeah, I feel like they're less—they have a smaller specturm of what is correct, compared to what I had at the start of the semester.
I [32:18]: And where is—why do they have, um, from where does that smaller spectrum come? Like, why do the instructors have that smaller specturm of, you know, acceptable and unacceptable?
S [32:30]: Uh, it could be a combination of things. One, again, I have just being a student more sympathy toward students, so I might think something thats, "uh it's pretty correct, but not hundred percent", so I'd say that's fine. And also more, um, moreso instructors I think... they just have a deeper understanding of the topic.
I [32:51]: Mm-hmm.
S [32:52]: With experience in physics, you have a much more clear-cut idea of like, "yeah that's correct. That's what Newton's Second Law means." Versus my more introductory understanding of the topic. Which is why also grading it was—why grading other people's assignments was kinda weird, like, if I've just learned the content myself, it's hard for me to know exactly that they used a correct discussion of, uh, the topics at hand.
I [33:18]: Yeah, let's talk about that for a little bit. You describe that—just the entire peer review process, especially at the beginning, as being a bit weird.
S [33:28]: Yeah.

I [33:29]: What do you think—let's just talk broadly about your perspective on the validity of using peer review, as a way of generating grades.
S [33:39]: So... in my personal opinion, I'm, I'm not a fan of it.
I [33:44]: Mm-hmm.
S [33:45]: I think that, um, there's too much—I think that the system of having TA—a TA or a professor grade it works better just because [glitch] whose internal compass of what's correct varies much less than the broad spectrum of students.
S [34:03]: And then on top of that, I think instructors have a more, again, with their experience like they're paid to know this material. Like, we're just learning it. I don't think it's exactly a great idea to have people without strong understandings of the topics like, strong proven understandings of the topic creating these papers. And I also, personally, didn't find... find it appealing that the grading assignments were weighted so heavily. Um, like, we found that they're super comparable in weight to the actual lab itself, which I found was like a lot more work. And a lot more—my effort was put into the actual lab, and I found the grading a little bit more arbitrary, in my opinion.
S [34:50]: Um. That's just my personal opinion.
I [34:54]: But beyond just the, um... beyond the matter of the peer assessment system actually assigning grades which get counted, what's your attitude toward the aspects of peer assessment? The fact that you are watching other um, so you used peer assessment to assign a grade, but in the process of doing so you also watch a lot of your student—your fellow students' things. Uh—
S [35:24]: [inaudible]
I [35:24]: Pardon?
S [35:24]: You mean, like, it's merits as a learning tool?
I [35:28]: Yes.
S [35:29]: Um... so, I can see the argument that uh, the repeated exposure to videos on similar topics like, reinforces the concepts that they're about. Um, but, I think that's a dangerous game to be playing when, one, the people who submitted the videos could be wrong and you'd be reinforcing the wrong concepts, and, two, um... uh, less on it just reinforcing the wrong information more on just its effectiveness as a leading strategy... I found that when I was watching the videos, uh, it was a much more disassociated view of the topic. I wasn't really looking at... the actual physics discussion, moreso the fact that they are talking about what they're supposed to be talking about.
I [36:20]: Hmm.
S [36:20]: It's a bit hard for me to put into words, it's... I was... mentally not focused on the actual material. I was focused on... moreso, how the material is presented. And um, how correctly it was presented, rather than, um... like, the actual learning of what Newton's second law, or third law is.
I [36:42]: And you describe the focus on the presentation of the concept rather than the concept per se as—
S [36:50]: Yeah.
I [36:52]: —dangerous, or perhaps, setting a bad precedent? Could you expand on that a little bit?
S [36:59]: Well, the dangerous part was more in reference to, uh, if the presenter of the video you're watching is wrong, and you don't have super strong understanding, and you're watching the video, you're—you're watching the video and reinforce the concept "okay, that's what Newton's Second law" even though that's not what it may actually be. That was more the dangerous thing. For me, the focusing on the uh, content presentation as opposed to... focusing on the presentation as opposed to content was more of my uh, reasoning for... why it isn't super effective even if the material is correct.
I [37:34]: And super-effective at doing what?
S [37:38]: Hmm?
I [37:38]: Super effective at doing... what.
S [37:42]: Reinforcing the material that we're learning.
I [37:43]: Okay. Um... uh, can you talk for a bit about the role of uh... the role of communication itself in the practice of science.
S [37:58]: Um... I feel like the role of communication is pretty important in science. With, um, a field that's so focused on research and data, um, it's incredibly important for communication to be clear and concise and unbiased, uh, because the way you frame any topic or issue uh, can totally change the way it's viewed. Like, just saying—just saying twenty percent as opposed to a fifth. Um... can bring in different connotations, right?
I [38:31]: Mm-hmm.
S [38:31]: Like even if you're using actual math. So, I think that... in the science community, the like... focusing on keeping dialog open but still uh, peer reviewed, and still uh, accurate... but... easily—easily distributed, is like a pretty big issue.
I [38:52]: Mm-hmm.
S [38:53]: And, uh, one of the goals the instructors had in implementing these peer assessment activities, uh, was to try to introduce um, experience in communication and in evaluating communication in an introductory classroom. Can you tell me a bit about what you think about the successes and failures of this program?
S [39:14]: Um... so, I didn't feel that when we were... when we were doing the uh, so when we were doing the lab videos, uh we had the weekly meetings where one of the meetings we had to have like an intro or the code done, and then, have to start on our Tracker data—gathering our data. And we met—we reflected on that, we moved on, made the video, and turned it in, and it was—it was next week. Uh... so the way the class is structured, I didn't feel was super effective at um... analyzing the actual how well your video communicated what it wanted, because when we were discussing it among our peer groups, uh... our videos weren't done, they weren't packaged in the way they were going to be presented. It was a more rough form of this is what—this is what I have so far. Um... but if you're talking more of just the assignments themselves, of like peer evaluation, and like how those affected like... like my learning of like how to analyze communication... um, I think it did a good job of giving me like a, uh, a general barometer of what a good narration looks like. Um, and what good content presentation can look like.
S [40:31]: Um... however, I don't know that it did... the course did a great job of defining that early on. Cuz it was more, it was moreso you just kinda figured it out based on how the instructors graded it, and because, like, people's presentation [chops?] varied so greatly, it's kinda hard to get like an objective grade like "oh, this was presented well."
I [40:55]: Mm-hmm.
S [40:55]: So... just, in my opinion it did a decent job, but it could be better.
I [41:00]: Alright, um, I'd like to talk about something a little more specific to the exercise we just did. Uh, when you were writing down your ratings, at the very end you said you went back and you did a "sanity check". What was going on there?
S [41:16]: Um, so, when I'm grading these, uh, I look at the rubric as a like, so, I think about the video as a whole, as like how good is that video in general. Um, my sanity check is just, okay, going through these, these grades... do these exp—do these match up with how I felt about the video overall? And if so, why—like, why are they different, should my opinion of the video itself be reevaluated, or was I being too strict or not strict enough on this specific subsection of gradings.
I [41:55]: Mm-hmm, Uh, and, uh...
S [41:58]: And for me the sanity check actually comes from, um, when we were doing the gradings uh, for the actual lab, uh calibrations, it said was this video better than yours or worse than yours. Or, like, much better or worse than yours, or about the same, that's where that comes from.
I [42:17]: Hmm. Uh, wonderful, thank you very much. So I think in the last quarter hour here, let's move on to the final stage of the lab. The ball rolling example video that I emailed you, is some observational data of a ball rolling—you saw it at the first interview—followed by a tracker screen where you get a position vs. time plot. What I'd like you to do is to watch the video, and then explain to me, um... what the physics is going on, and in particular, why the tracker plot has the shape that it does.
S [42:51]: Okay. Ball roll example... [operating computer]
I [42:56]: And if you could please screenshare so I can follow along with you.
S [43:00]: Yeah, perfect. Lemme uh... [operating computer] [muttering] pull this up for a sec... screenshare...
S [43:19]: Alright. So the ball's starting initially at rest, it's at the left side of the frame, so that's where the origin's going to be as the start point. Positive x is the right—positive—to the right is the positive x direction. Um, relatively constant motion, I'm expecting a pretty flat—uh, a pretty straight line for the tracker data. Uh, cuz the ball didn't really increase or decrease in speed too much, but we'll see—
S [43:47]: And yes, as I predicted the line is relatively straight. The slope, um, is pretty constant. Um, which means that uh, because the derivative of th eline is its slope, the derivative of position over time is velocity over time, uh, which we learned earlier in the year. Um, and if that's not changing, it means that there is no unbalanced forces acting on it as it's rolling, it just has the initial force and it rolls constantly. So the role of friction is what's causing it to have that slight little bit of bend over here. It bows out just a slight bit, uh, over the middle here. Uh that's just cuz the ball isn't perfectly frictionless, neither is the ground, it's a bit dirty, the tennis ball's fuzzy so it's catching a bit.
S [44:36]: Um, but... it is still rolling at a relatively constant speed, uh, with no unbalanced forces acting on it after—after the ball's released.
I [44:45]: Okay, and uh no—no unbalanced forces acting on it?
S [44:49]: Uh, no unbalanced forces acting on it after—uh I mean, liter—actually friction is acting on it. Friction is acting on it in the negative x direction. But um, not—relatively negligibly. In the sense that, like, what our videos are asking us to do. Show a video of constant motion—we couldn't get like actually constant, but this is about as good as it gets.
I [45:18]: Very good. Now, initially, you predicted that the origin would be placed at the starting position of the ball. [But in the tracker data it's not...]
S [45:26]: [Yeah. It's—uh—[inaudible]]
I [45:27]: Yeah, why is there a difference there? Why do you think...
S [45:32]: I didn't account for the fact that his release point, um, so... the ball takes a bit to speed up, so he probably chose that uh origin uh as a start for the data because that's probably around where the ball reached constant speed.
I [45:47]: Mm-hmm.
S [45:49]: Because it has to accelerate. And he has, uh, to make sure his hand is gone from the—make sure his hand is gone from the frame, to show that like he fully released it. And that's probably why. Because I said that when I didn't see where his hand was gonna be.
I [46:02]: Okay, very good. Um, so, show me—tell me a little bit more about the role of friction in producing the curvature of the position vs time plot. How could you change the curvature? What could you do with friction to change the curvature of the plot?
S [46:17]: Oh—uh—sorry—should I, uh, screenshare...
I [46:23]: Uh yeah, if you're going to refer back to the video, please screenshare.
S [46:24]: Yeah cool.
S [46:26]: So you're saying, how could I change the curvature of the line in tracker?
I [46:31]: Yeah, specifically with respect to friction.
S [46:36]: Oh. Huh [plays lab video]
I [46:36]: No, I meant the ball—the ball rolling example.
S [46:39]: Oh yeah, it just got switched.
I [46:42]: Oh that's fine.
S [46:43]: Perfect. And let me make sure [mumbles]. Ball roll example...
S [46:53]: Alright. Um, so... friction is the force that acts in the negative x direction. Um, like on an atomic level, even smooth surfaces are rough, atomically. And so, the atoms rubbing up against each other, catching up on each other, are causing resistance in the direction—resistance against the direction of movement. And so to uh, flatten this line a bit more, uh, what could be done are, uh... like an actually robotic release for the ball, where you know it's gonna be constant speed when it's rolled. Um, on top of which, using a smoother ball, such as like a cue ball from like a pool table. Or something like that. And a smoother surface, uh, so an air-hockey table is super smooth—something like glass or, uh, like, polyurethane covered wood. Something that's more perfectly—not actually perfectly, but more perfectly—more flat than this kind of dirty porch.
I [47:59]: Okay. Very good.
S [48:00]: Those things would change the line, the slope would be more constant.
I [48:03]: So if you were drawing a free-body diagram of this ball, as it was rolling—let's say at that point in the video, what would the forces you draw on it be?
S [48:14]: Normal force pointing upwards, uh, from the ground up on the ball. Force of gravity pointing downwards. They're equivalent though, in the line size—in the arrow size, because uh... in the y-direction, because no outpu—no unbalanced force in the y direction means no movement in either way. Um, And then you'd have uh, to the left, you'd have friction acting—you'd have friction acting on the ball.
S [48:42]: Um, if you drew the free-body diagram from earlier, like over where his hand is still touching it—but he actually has that completely cut out, so you wouldn't have any free body diagram in which you have force applied.

I [48:55]: Now, when you were evaluating the lab video at the beginning of the hour, you mentioned drag, and you mentioned that drag in this case was negligible. Can you tell me a little bit about how that relates to this situation here?
S [49:07]: Uh, again drag is pretty negligible. Um, it... it's more of a... it's... it's the idea that uh, the fluid that an object—an object is moving through is resisting its motion in the opposite direction. Um, because like air, air for example is going to get pushed to get out of the way. And it's pushed—and usually it's used to describe the phenomenon of an object falling and reaching terminal velocity when the drag force is equal to the uh... force of gravity. It stops accelerating. Um, cuz there's—it's not resisting it anymore, the object stop—uh, the force of drag is equal to the force of gravity. Um... but in this situation, the object's not moving too fast, it's not catching a ton of air because it's a sphere, um, so... uh—it's not playing much of a role. Cuz drag is related heavily to velocity as well.
I [50:01]: [Very good]
S [50:02]: [And the drag] coefficient of [mumbles]
I [50:06]: And so finally I'd like to discuss a bit about one of the basic premises of this sort of investigation. Uh, the position versus time plot there is the position versus time of that point that the experimenter has been clicking on in the middle of the ball. Um, but of course not every—not every molecule of the ball is moving at that same—at that same velocity, because it's rolling, it's undergoing actually a fairly complicated sort of motion.
S [50:35]: Okay.
I [50:36]: So talk about how we're justified in describing the physics of this system in terms of the motion of, uh, this selected point.
S [50:45]: Um... so... uh, the system's basically when we're analyzing... analyzing it, we're analyzing it essentially as a point mass. So we're reducing all of the cluster of, to basically the center of the object. And just trying for the—trying for the most accurate click each time. So that's—also probably played a role in the slight bowing of the graph, is he's not going to get the exact center of mass of the ball every time he clicks on it.
I [51:13]: M-hmm.
S [51:15]: But it's a general approximation of the system, um, that the click on the center, it's not going to be obviously the actual assortment of atoms, how they're moving every time, the exact...
I [51:26]: And uh, so tell me a little bit more about how uh, failing to correctly identify the center of the ball might yield—might help to yield the curvature of that plot.
S [51:38]: Um... so essentially, the way the tracker data is plotting, um, this graph, it's taking the scale that you put into it, uh, and, uh just... uh scaling it to your point clicks relative to the origin that you select.
I [51:55]: Mm-hmm.
S [51:56]: Um. So if you click somewhere that the ball isn't actually moving, uh, the data's still gonna look like the ball went there, even though you clicked incorrectly. Uh, so if you click too far to the right, for example, it could mistakenly be seen, or too far to the left for example, in this case, it could be mistakenly be seen as the ball slowing down as opposed to just some misclick on the tracker data.
I [52:24]: Okay, very good, um, and are there any other sources you can think of which might contribute to the curvature of the plot there?
S [52:33]: Curvature of the plot. Other than user error of Tracker, and friction?
I [52:38]: Yeah, those are the two you've identified so far.
S [52:41]: And other than... obviously drag is minimal, um... thinking...
I [52:52]: Uh, "no" is an acceptable answer.
S [52:54]: What? Oh! No, I don't believe so.
I [52:58]: Um. Very cool. Alright, thank you so much, thank you so much [name], I'm going to call the interview here.
[logistics etc.]

F.14 Student E, Summer 2017 (Pre)

[logistics etc.]
I [02:15]: So let's begin with the practice physics problem. Now, I'm gonna ask you a question, I'm gonna want you to think aloud while you're trying to solve the problem. While you're doing that I'm not really going to be askign questions, but if you're a little too quiet for too long, I might prompt you to keep talking. And that's about all that I'll say.
S [02:35]: Okay.
I [02:36]: Okay, so. This is a two-part estimation question. One, what is the volume of the room that you're sitting in right now, and two, if you were to turn on an ordinary kitchen faucet in that room, about how long would it take for the room to fill up completely with water?
S [03:00]: So like um, an [EMED?] problem, right?
S [03:06]: So you said what is the volume of the room.
I [03:08]: Yeah. The room that you are in right now.
S [03:12]: Okay, so, I have... I have a crate over there that's... maybe a foot long. So maybe a foot and a half. Like eighteen inches. [writes]
S [03:38]: [Looks like?] it could fit... [gestures]
I [03:44]: Keep talking please.
S [03:46]: I'm just trying to figure out how many... crates would fit along the wall. [counts under breath] three... four... five... six... seven...
S [04:09]: Okay. Probably... nine crates. Nine crates go along this wall here [gestures]. Um... then... that would be the same on the other wall. The other two on this side are smaller, I mean. Smaller than the other wall. Um, I could probably fit about two... three... four... mmmm... maybe six. Six crates on those walls. And for the height of the room, hmm... I'm five foot one. I'm pretty sure the shh... probably an eight foot ceiling. So if I can fit nine crates along one wall, and I said they were eighteen inches... one hundred and sixty two inches, divide by twelve, thirteen and a half feet.
S [05:47]: Then six... times eighteen inches. Divided by twelve. That's nine feet. So thirteen and a half by... nine by, what'd I say? Eight? Nine hundred and seventy two feet. Cubed. Here. [Sure?]
S [06:17]: And then what was the second part of the problem?
I [06:21]: If there were an ordinary kitchen faucet in that room and you turned it all the way on, about how long would it take for the room to fill completely with water?
S [06:29]: Okay, so an ordinary kitchen faucet, let me say it fills... one cup of water in... um... forty five seconds. And a cup of water is... um... say... I'll make this easier.
S [06:53]: Um... two... trying to fit in... maybe two, like, cups. Measuring cups. Um... so... so for those two cups of water and forty-five seconds, um that would be... how long would it take to fill...? oh shoot, we don't have any of this in feet. Instead, maybe... let's say... one glass of water is... maybe four inches. By... three inches. By three inches. Well, the glasses that I have at my house. Three by three, that's how many inches... thirty-six inches cubed. And let's say that's done in forty-five seconds, then it would take... nine hundred and seventy-two feet cubed... [writing]
S [08:17]: Um... if there, how many inches cubed are in one foot cubed. One cubic foot... is twelve inches by twelve inches by twelve inches. So... cubed. That's seven—seventeen twenty eight. Okay.
S [08:47]: So. We're saying how many? Thirty six, have to do seventeen twenty eight divided by thirty six, that's forty eight... and that's forty five seconds. That's two thousand one hundred and sixty seconds, divide that by sixty to get minutes. About thirty-six minutes.
I [09:17]: Okay, very good, thank you, that's can you just summarize your final estimation?
S [09:25]: So I estimated that the room that I'm in would be nine hundred and seventy two feet cubed, and because of that, um... wait... no no no, I did this wrong. I did this wrong. Hold on. We... oh, okay, if there's seventeen twenty eight inches cubed in one foot cubed, then we need to do nine hundred seventy two times seventeen twenty eight. And then divide that by thirty six, multiply by forty-five seconds, divide by sixty minutes, divide that by sixty to get hours. [laughs] Okay, that's a lot longer of a time. So, I estimated nine hundred seventy two feet cubed, um, and then I said that if I could fill—if the sink could fill a cup in forty-five seconds, and my cup of water is four inches by three inches by three inches, um, then... that was thirty-six inches cubed. Um... and then... I figured out how many inches cubed were in a foot—a foot cubed. And used that to figure out how many inches cubed were in nine hundred seventy two feet cubed, and that was a really large number.
S [11:12]: Um, divided that by the thirty-six inches cubed that were in a cup, er, the cup that I was using. Multiplied that by forty-five seconds, divided it by sixty in order to get minutes, then divided it again by sixty to get hours. And that was a hundred and eighty-three point two hours.
I [11:36]: Okay, very good. And what was is it—you were about to offer a very different answer than that, what was it that made you think you'd done it wrong?
S [11:39]: It was because I was using the number of inches cubed that was in one foot cubed, instead of the number that was in nine hundred seventy two feet cubed. So, I forgot to... do the... um, dimensional analysis that I started. [laughs]
I [11:56]: So that is the error that you made, what do you think it was that prompted you to realize that you had made an error?
S [12:02]: Um, because thirty-six minutes is not a lot of time to fill a whole room [laughs].
I [12:08]: Okay, very good. So that is generally how, um the think-aloud study is going to work. I think you've got a good grasp on it, to the point where we can move on to the next part of the interview. So. If you go to the email that I sent you about an hour ago, uh, there is a lab video that begins with a... jar of peanut butter in Tracker rolling across the screen. Um. I want you to bring up that video and also I've—for your your conven—conv—ehh. For your convenience I've attached the usual rubric. Um, and while you're evaluating the video, I would actually like to see what is on your computer screen. Uh, so please clear any personally identifying information, or any sensitive information from your desktop, and then share your screen with me, please.
S [12:59]: Okay. Let's see if I can get this video up here.
I [13:15]: It might take a moment to download.
S [13:18]: Oh. Alright. So... [sighs]
S [13:40]: Okay... so you wanna see the video and the rubric?
I [13:46]: Uh, yes, I want to see what's on your desktop as you're watching and evaluating the video.
S [13:54]: Okay, I have the rubric in the... picture, I'm going to put it on... like, a word document.
I [14:04]: Okay. And yes, when you're filling out the rubric, unfortunately you can't type onto the picture itself.
S [14:12]: Yeah.
I [14:13]: So please just put it in a—put your ratings and comments in a text file, or on a piece of paper or something.
I [14:21]: Okay.
S [14:24]: [muttering] instruction... content models...
I [14:44]: While you're getting that set up, I do have one quick question. Uh, when I first stated the estimation question, you remarked kind of under your breath that it's, it's like a BMAT problem, or an EMAT problem? What is the name of that test, and why did it sound similar?
S [15:02]: Oh, I said BMED problem. I just took BMED 2210, and we... did a ton of estimation problems just like that. Um, every day [laughs] so... um. That's why I said that.
I [15:21]: BMED is for biomedical engineering?
S [15:26]: Yeah.
I [15:27]: Okay.
I [15:35]: In physics, we sometimes call those "Fermi" problems, for Enrico Fermi. He liked to ask a bunch of, you know, estimation questions.
S [15:56]: [muttering] where is the screenshare? Here it is. Okay.
S [15:57]: So...
I [15:58]: Okay, very good.
S [16:00]: Yep.
I [16:00]: Alright so please, uh, give it a watch and evaluate it in the way you normally would. Um, I understand that when you're watching and listening to a video, it's difficult to speak and think aloud. But when you are filling out your rubric, or stopping and making decisions, please think aloud.
S [16:18]: Okay.
I [16:19]: Alright.
S [16:33]: Okay. Well there wasn't a whole of an introduction. [under breath] introduction. [sighs] Lacks... fundamental principles, and introduction of New... Newton's

second law. And momentum principle. And terminal... velocity. Is not sss... used [inaudible] Tracker for vpython. And—okay.
S [17:54]: Oh. Which... this is lab one?
I [17:59]: Yes, that is a lab one video.
S [18:03]: Okay. Well in that case we don't need to talk about terminal velocity. We can talk about Newton's Second... nope, we need to talk about Newton's first law.
I [18:15]: So it would actually help me a bit if you spoke a little more clearly.
S [18:18]: Oh, sorry. I said, in that case we don't need to talk about terminal velocity or Newton's Second Law, but we need to talk about Newton's first law and kinematics. Um... and it... doesn't discuss the use of Tracker or VPython. There just wasn't much of an introduction at all. Okay.
S [19:42]: [under breath] constant velocity.
S [19:45]: Okay, let's take a look at the rubric [for a second?]. Um... alrighty. Well—oh. Let me plug in my computer.
I [20:03]: Okay.
S [20:13]: Well, production delivery. He speaks clearly. And slowly... so the audience can understand. That is good. Um... content models... starts with... Tracker... video... and explains... [he?] shows that the peanut butter jar moves at constant velocity by showing the tracker... graph.
S [21:08]: Okay, let's continue [sighs]
S [22:14]: Okay, well that is not correct, because if you're calculating the change in velocity over the change in time, then that would be acceleration and not... velocity. To find your initial velocity... I mean you start from rest. So it should be zero. So, when discussing the... tracker video [mutters] initial velocity [inaudible] calculated by... taking the change in velocity over the change in time. Even though practically, he took the change in position over the change in time.
S [23:32]: Either way, initial velocity should be zero, if starting from rest.
S [23:55]: Okay, let's get back to the video.
S [24:43]: Okay, that was a little patchy. Um... when discussing... the what if question, about the what does it mean question I think, [if?] I remember correctly. Question... about... forces, he... said that the individual forces could not be determined because they are quote invisible. However, we know that the peanut butter jar is affected by forces such as gravity, and friction. And we cannot determine their magnitude based on the data for [all of?] this lab.
S [26:51]: That was good! He used vpython to introduce Newton's second law... [go back?] a little bit [rewinds video] oh, that's what he did. Okay. Mentioned... constant... velocity was caused by a zero net force. Showed... kinematics... conclusion.
I [29:12]: I'm sorry, are you still there [name]?
S [34:38]: Hello!
I [34:38]: Hello!
S [34:39]: Sorry, I didn't realize it cut out.
I [34:44]: Oh, okay, that's alright. Were you still able to record you responses and ratings to the video?
S [34:52]: Um, you mean like, in my word document?
I [34:55]: Yes.
S [34:56]: Okay. Um. Here, do you want me to screen share it with you right now?
I [35:00]: Yes please.
S [35:10]: Okay. So... I'm not really sure where it cut out. So do you want me to just explain everything.
I [35:21]: Uh, y... actually, not yet. We can go over that explanation in a bit. Um... well... lemme think. No, no, no—yeah, if you could please explain them now, that'd be great. So organizational structure, I said um... it was just good. There uh, it wasn't horrible, but the introduction—actually I would probably put that at fair. Because the introduction was lacking in um, physics principles. Which is kind of the main point of the introduction. And he didn't introduce Tracker or vpython so I would say he didn't really fulfull what he needed to do.
S [36:12]: The content models, um, I—starts with—tracker video shows peanut butter jar, moves at constant velocity—tracker graph... Um... he used vpython to introduce Newton's second law, and that was probably halfway or more than halfway through the video. And that was pretty much the whole point of the lab. So... it was good that he... did do that, but he—then the only time he, um, mentioned that velocity—constant velocity was caused by zero net force or the relationship between those two things was in that code. So I would say that the content models were good.
S [36:54]: Content prediction discussion. Um, when comparing the observed data and model that he used... reasons for discrepancies between the graphs. Um, that was good, he... said that um... uh, he said that... the... that friction was the reason for, um, the peanut butter jar slowing down, which was correct. He didn't really say much more than that, really, in terms of Newton's second law. Um, other than fnet was zero.
S [37:31]: When discussing the what does it mean question about forces, he said the individual forces could not be determined because they're invisible. We jnow what the forces are, because we've learned about them. Um, and he even said that it slows down because of friction, but I think what he meant to say was that we can't determine their magnitudes based on the data from this graph.
S [37:57]: Um, and then also when discussing the tracker video, he claimed the initial velocity was calculated by the change in velocity over the change in time, even though what he told us to do was to take the change in position over the change in time. Either way—the—uh initial velocity would have been zero if it's starting from rest. Or—he didn't want it to start from rest, then...
I [38:23]: [sneezes] Excuse me.
S [38:25]: Bless you .
S [38:25]: Um, and he also didn't exp—well he said that the axis—the peanut butter jar was moving in the positive x direction, so I suppose that kind of explains why, um, he used vectors with only x components, but he didn't say it.
S [38:47]: The production delivery I put very good because he was, he spoke clearly, slowly, easy for us to understand. Um... didn't have any problems with the video. Uh... and... that is about it.
I [39:03]: Okay excellent, so keep the screenshare up actually, because what I'm going to do—uh so let's—this the set of ratings and comments that you would submit to webassign?
S [39:15]: Mm-hmm.
I [39:17]: Okay. So now I'm going to email you the expert evaluations for that same video. And I'd like you to look over them and compare them with yours as you normally would.
S [39:30]: Well. This is better than I usually get [laughs]. Um, usually I don't match up very well. So... where—said fair, no results in intro... no preview of major section, good description of lab's purpose, [inaudible] what if question should have been signposted. I don't really know what that means. Um... okay. What did I say? I said fair, okay. That's good.
S [39:57]: And... I... no explanation of physics principles or connection between principles... and that was for content models. Okay. I definitely see where they're coming from. I was... probably trying to... throw them a bone. But... I... don't understand why they would've given him a poor on models instead of on prediction and discussion, because I feel like the, um... physics principle—like the—explanation should be... um... under the discussion tab.
S [40:41]: Okay. Good. Very good observation model comparison, with good account of role of friction [inaudible] Okay. The account of initial conditions calculation was buried in a disorganized explanation of the python code. I agree with that. Um... That's also pretty much why I gave him a good. Um...
S [41:01]: Good—inadequate—um... what does it mean explanation. No motion net force connection. Good what if question. Physics principles appear to be soundly understood but not always clearly expressed. I think that in their um rubric, they—if you don't express it, you can't assume that they understand it. So... um... but I guess if you have a good what if question, and an—inadequate what do you, what does it mean question, then you meet in the middle. That's also what I put.
S [41:41]: Very good. Very good production quality, except for hard to read walls of code. I didn't find them that hard to read. But I suppose, you know, it could have been clearer, or better. Um... but... I guess we were... fairly similar. So that's good.
I [42:00]: Okay. Excellent, so now if I could see your face.
S [42:05]: Uh. Yes. Stop screensharing...
I [42:08]: Wonderful okay, so now I would like to go over a few followup question I have for this exercise. Uh, and I'm keeping aware of the time, uh, because we cut out for a bit, would it be possible to go until... uh, let's see, we started—we actually started at about five till eleven. Could we go to five past one?
S [42:31]: Sure.
I [42:31]: Okay. So firstly, you remarked uh, when you first saw the expert evaluations, "oh I usually don't do this close". Uh, could you tell me about that?
S [42:41]: Normally, I evaluate the um... the what are those, the calibration videos. And... they're evaluation are nothing like mine. I... I feel like they're inconsistent with what they rate harshly, and what they don't rate harshly. Um... and... I don't find, um, I don't find it easy to match what they say [laughs]. So, I just don't really know what to do about that [laughs] yeah.
I [43:18]: So does it seem, um, so you say there's an inconsistency with what they grade harshly, and what they don't grade harshly. Is the problem that it's inconsistent, or is the problem that there's too much harshness, would you say?
S [43:35]: I think the problem is they don't have an individual rubric for each lab. Because if you don't specify what we need in every lab, like—we have a general description in the instructions on webassign. Um and then we have this rubric that we're supposed to use for every lab. But they don't say well, we want you to discuss... um... I—I dunno. In this lab, we want you to... keep in mind um, constant velocity in the introduction or—let's try and give an example of what—how they graded [inaudible] Like... what they want emphasized in each lab, I suppose. Um... because—you're—you would grade more harshly something that's more significant in the rubric, so if they missed a major part of something that's in the rubric, um, you would rate it more harshly, I suppose. And, but if you don't have an individual rubric for each lab, we don't know what we're supposed to be looking for, because if we submit our own lab that we think is great, and they don't, how are we supposed to judge other people's labs? [inaudible]
I [45:09]: So, uh, if you had to um.. describe your general attitude toward the overall reliability of the peer evaluation system—not just the expert ratings for the training and calibration videos, but the ratings that you give each other, that you peers—that your give your peers videos, and your peers give you—what's your overall attitude toward how reliable a system that is?
S [45:34]: Um, I, I don't really think it's reliable. I think it's a... I think it helps us understand what, [glitch] in our own video, by looking at other people's videos. I think that's a good idea. But... I don't think it's... reliable to use other people's... rating systems, just because we're all like, none of us are... not that we're not any more qualified than each other, but, after we turn in our own lab, we haven't been given any... feedback. So we don't really have anything more to base it off of.
I [46:22]: So by the time that you're writing your lab two video, you haven't yet seen your feedback from your lab one video.
S [46:31]: Yeah.
I [46:32]: Yeah, that is a significant complaint, I can see that. Um...
S [46:36]: Yeah.
I [46:38]: I'd like to walk back a little bit to some—to something I notice you did when you were doing your evaluation. While you were reciting your evaluations to me, after we got interrupted, you changed your rating for the first item from good to fair. Uh, could you explain why you did that.
S [46:59]: Yeah, um, I was looking at what I—the comments I wrote, and then I looked at the rubric, and realized that... it really just... didn't deserve a good. I, it's, um... main part of the first section of the rubric was the introduction, and the introduction was lacking in almost every aspect, so... just deserved a fair, like [laughs].
I [47:29]: And for—and on another item, as you were describing it, you said that you, you felt like throwing him a bone. Can you tell me about that.
S [47:40]: Yeah... uh, he did a fine um, job presenting each of his models. I don't think he had any problem with his models. But that, I mean another thing with the rubric. They graded um... I'm gonna look at the way they did their rubric. Okay they give him a poor on the uh... models. Um, section. And I would have—if I was gonna give them a poor I would have done it on the discussion section, because I think that he presented his models fine, but he didn't discuss them properly. And they said he... he discussed it fine, but they didn't, he didn't present it well. And so I think that's a little bit vague.
I [48:29]: Mm-hmm.
S [48:32]: Um. So I guess... he did a good job.

I [48:36]: Uh. Okay. And would you say uh, how would you describe... the sort of decision making process you go through, when you decide to give a poor grade, or a very high grade. Do you find yourself inclined more toward one than the other?
S [48:59]: Um... a poor grade a very high grade? What do you mean, like...?
I [49:08]: Well, when you said like you were trying to throw him a bone, it seemed like you were referring to some sort of internal inclination to give people the benefit of the doubt, or maybe lean toward higher grades. And I just wanna explore that a little bit.
S [49:20]: Well, I guess... I would rather... not that I would rather give a high grade, but if—I think somebody's on the fence between poor and fair, I would probably re-evaluate and see if they... seem like they knew what they were doing. Like, in here it said, physics principles appear to be soundly understood but not always expressed. They—I... probably threw him a bone because it said appear to be soundly understood, whereas in reality I guess when you're grading if you don't express them correctly, then you can't assume that they're soundly understood.
I [50:06]: Mm-hmm.
S [50:08]: Saying... [inaudible]
I [50:13]: And on that note, you mentioned something about, um... talking about whether y'all were qualified to grade these videos. I'd like to talk a bit about what it means to be qualified, or authorized to give grades to videos.
S [50:31]: Well, we are asked to grade these videos right after we turn in our own video. We aren't given any feedback on our own video, um, before we grade other people's. So... I feel like... like if we all think our own video is good, then... and obviously they're not all the same level, then we are just grading everyone's videos based on our own knowledge, and if somebody doesn't understand the physics principles, then there's no way they can... grade somebody else's, if they don't know what to look for.
I [51:17]: Are you worried that's an endemic problem among your peers? Are there peers out there that don't understand the physics principles?
S [51:25]: No, I just think it's—I mean I don't know, maybe. But um... they... I just think we should be given feedback, um, or more specific rubric, so we can look—so we know what to look for in somebody else's, um, video. Like, so we know what to grade harshly and what not. What's most important when we're looking in a video.
I [51:52]: Mm-hmm. So out of... what do you think—so the instructors of this course, uh, they implemented this set of video lab exercises with the peer assessment, what do you think their goals were?
S [52:16]: I mean, I could see how us grading each other's video will help us better, number one better undsertand what the lab was about, [do it?] and forget about it, um number two, help us to make a better lab report next time. That my... lab two video was a lot better than my lab one video. Um... after looking at everyone's. I think it helps people personally, but I don't think it helps the person that is being graded.
I [52:50]: I see. um, and why—and it doesn't help the person who's being graded so much, because... what is it that's missing?
S [53:05]: Um, just like—sorry you cut out for a second, what's missing from the person that's being graded?
I [53:12]: Yeah what's missing, do you think, that would help the person who is being graded.
S [53:17]: Just like what I was saying before, if we were... like, we have no room to judge other people's videos at the... at this point, after turning in our own video. Like it's more fair... I mean, i don't know how many people grade each video. Is it maybe... [glitch] grade each video and then the scores are averaged? I'm not really sure. But it's more fair if everyone's graded by the same person, or just...
I [53:53]: Okay, so there is—if you go on Piazza there's a document that explains the whole peer evaluation process, how videos get assigned um, how many people are evaluating each video, it's three other students who are evaluating each of your videos. Um, but, what—what sort of qualities would enable you to give better grades, you know, would qualify you to give more accurate grades to your peers?
S [54:24]: Um... like the reason I said a TA before is because they have a deeper understanding of these physics principles. This is all of our first time. Maybe not first time, but we are in physics one. And we haven't done anything too, um... that, uh... gives us that understanding of... physics principles that we can recognize in other people's videos, if... [coughs] excuse me. If we may or may not be able to recognize them in our own video.
I [55:04]: Okay. Well very good, so we have one last part of the interview, this has been a very illuminating interview so far. If you could please go back to that email I sent you, there's a very short video clip of a ball rolling across a table.
S [55:21]: Ooo-kay.
I [55:23]: I believe it's called ball roll example dot m-o-v.
S [55:28]: Alright.
S [55:39]: Okay, should I just...
I [55:41]: Ah yes, screenshare please. Uh, I'd like you to watch it and then, uh construct for me an explanation of the physics going on with the ball rolling, and uh, an explanation of the shape of the tracker plot.
I [55:56]: Alright... [share?]... okay, [here it goes?]
S [56:25]: Okay. Well, is it gonna show—okay here it is.
S [56:41]: Well, the ball rolling, this is basically lab one. The ball rolling was I—I dunno if they meant for it to be at constant velocity, I suppose they tried. There's a little bit of an arch here, um, which means the velocity wasn't... constant, but I guess it's hard to control. And, like we said in lab one, because there's... constant velocity, there's no net force, so the ball will just continue... to roll. Um. I'm not sure what, what... you want me to do... [laughs]
I [57:36]: Okay, so you say you want—or that you intend to get constant velocity, but that it's quote hard to control.
S [57:45]: Yeah.
I [57:46]: What do you mean by that?
S [57:46]: Well if you're just pushing a ball, you aren't... [glitch] [force?] that you put on this ball. And you can't control—I can't tell what surface they're on. You can't really control the, um... friction.
I [58:03]: Mm-hmm.
S [58:03]: It's... creating I guess, they're using a tennis ball, so maybe something that, um didn't have as much texture to it. Uh, but that's getting a little—like—getting a little bit technical. Like, the arch is very slight. And it... if you did a model, like a computer model for this, it would almost match. Um... for constant velocity.
I [58:34]: Can you tell me a little bit more about the role of the, uh, the push of the ball? Cuz you said it was hard to know exactly what that force is, or hard to control that force. Tell me a little bit more about the role of the push on the ball.
S [58:53]: I mean, that's the initial um... your initial force. It's like, Newton's... law said that... uh... object in motion will stay in motion unless acted upon by an outside force. And... that's your first force, and the ball's going—it's continuing with the same motion, um, by an outside force, and that was friction. Um... other than that, I mean we don't know the magnitude of that... but that's what... caused it to roll in a straight line at a constant velocity.
I [59:31]: Okay, and can you show me... can you tell me about the relationship of the force of the push to the shape of the plot specifically that's up there, in the top right corner?
S [59:48]: The force of the push to the shape of this graph here?
I [59:52]: Yeah because you brought up—I'm asking this because when you were describing the fact that the—that the position versus time plot has a little bit of an arc. Um, you were talking about the force of the push as being something that's kind of out of your conrtol or experience. And I want to know how the force of the push relates to the arc of the plot.
S [01:00.20]: Well... the force of the push... mmm... is relative to the velocity. Because veloci—like, final velocity is equal to initial velocity plus fnet over n delta t. And we want here, [for constant?] velocity we want our fnet to be zero. Um... but... it's not going to be exactly, excuse me exactly zero. Um... because this is just a practical experiment, and those um, those models are... idealized.
I [01:01:04]: Mm-hmm.
S [01:01.25]: Um. And... I mean... the slope of this... is the velocity, if I didn't say that already. Cuz it's the position time graph.
I [01:01:19]: Mm-hmm.
S [01:01.13]: Um. Yeah.
I [01:01:22]: And can you tell me a little bit more about the role of friction in, uh, in the arc of the plot.
S [01:01.26]: Friction is moving... a force that's pro—going against the motion of the ball. So it's going to... slow down the ball. Which means this is going level—well it would level out until it's not moving anymore. So, the... it won't be as steep a slope in the [following?].
I [01:02:01]: And what's causing the friction in this case?
S [01:02.11]: The friction is caused by, uh, the ball rolling against this probably cement, on... it... the, well it's just... the force... of... yeah. The ball hitting the cement. Slowing it down.
I [01:02:27]: So you might be able to control the amount of friction by lubricating, the surface, or choosing a different material to make the ball and the floor out of?
I [01:02:39]: Yes. That's why in um... physics problems, they always say, "oh, on an air-hockey table", or on a... "an ice-skater" or something because there's not a lot of friction in those cases. But a tennis ball has a lot of texture to it, and it looks like this... um, I guess it's rather smooth. But um, for something like a tennis ball, it might have been better to use something smoother like a bocce ball or something.
I [01:03:10]: And are there any other possible—can you think of any other explanations for why that plot is arcing downward?
S [01:03.18]: It's arcing downward, why it would be slowing down?
I [01:03:26]: Yeah.
S [01:03.26]: I mean... really, the main force going against the um... going against the initial push is friction, because it's only moving in the x—it's only taking into account the x-direction. I mean, maybe drag force from the air, but that's not... hugely significant.
I [01:03:58]: Why do you say it's not hugely significant? How do you know that?
S [01:04.26]: Um... because it's not... it's moving through the air, but... the friction force is going to be larger than the drag force. There... it's like in the second lab, when it said velocity, um... what is the... how—how would the motion be affected if it was thrown down versus if it was dropped.
I [01:04:30]: Mm-hmm.
S [01:04.08]: This would be like if it was thrown down, because you have a force going in the positive x-direction. And you're... being counteracted.
I [01:04:44]: Alright. Um... let me see if I have anything else, I actually—I think that's about all I have for you, we're coming up on one o'clock, I'll let you go on time.
[logistics, etc.]

F.15 Student E, Summer 2017 (Post)

[logistics, etc.]
I [04:18]: So. For our interview today, uh, we're going to essentially repeat uh the same protocol that we did last time. Uh, the basic idea being that I want to ask you more or less the same questions at the beginning of the semester and at the end, to see how your thinking has changed uh since the last time we spoke. Um, one difference though is that we're not going to start off with the think-aloud warm-up, because we've—we've been through that already and it's frankly not the most difficult skill to develop.
S [04:46]: Yeah.
I [04:46]: So, to begin today, uh, we're going to start with a... think-aloud protocol evaluation of a lab video, which I've included in that email. I'd like you to watch the video, uh, walk me through your reasoning, uh, about when you assign the ratings and comments that you do. I'd like you then to compare your ratings to the expert ratings. And then we're going to have a discussion about that experience.
S [05:13]: Okay.
I [05:14]: We'll end the interview like we did last time, with a bit of discussion about some observational data, which is also in that email. Um, followed up with a kind of a freeform discussion about physics principles, um, and related concepts.
I [05:28]: Uh, so please open the lab video that I included in that email, uh, and if you could please share your screen with me so that I can follow along with what you're watching.
S [06:00]: Is it just...
I [06:02]: It's, uh, it's a screen—the filename is a string of characters. The first frame should be a picture of some tracker data.

S [06:13]: Oh, I see.
S [06:28]: It's just taking a while to download.
I [06:31]: Yeah.
S [06:57]: Okay.
S [07:09]: Uh, is there a copy of the...
I [07:09]: The rubric? A copy of the rubric should have been included in that email. Lemme know if it's too small to see.
S [07:14]: Oh, I see it.
I [07:16]: Is that a full-resolution copy? Ah, no it isn't, it always does that. Uh, give me one moment and I will send you a... higher-resolution copy that's actually readable.
S [07:30]: Mm-hmm.
I [07:54]: Okay, I've sent it.
S [08:09]: Oh, perfect.
I [08:10]: Yes. If you could please type your ratings and comments just on a notepad document or something.
S [08:17]: Okay.
S [09:15]: [So?] first of all, talking about the introduction, so far, very limited... starts the tracker video without introducing the physics principles.
S [09:48]: Content models... content prediction discussion... content overall, and production delivery...
S [10:44]: Let's keep going.
S [11:09]: Okay. So he said that... [rewinds] he said that the assumption... we are making the assumption that fnet is equal to zero. But he did not explain that this is because we are rolling the jar at constant velocity. Which, again means that there was no explanation... of physics principles. Um, he started answering the what if questions, that goes under content overall.
S [12:31]: Addressed... what if... question... on axes. However... described... the axes as straight... which... which is irrelevant... because an axis... is already a line. He should be describing further direction... and why he chose his initial starting point.
S [13:49]: Okay. Okay. That's what we're talking about. He said that it moves in the positive x direction... okay, that's better. So... he mentions that if the axes... were flipped... as the question asks, then... the velocity... would not change in magnitude... but only direction.
S [16:25]: [before we?] go on to [get?] content models, um... so content models, we already saw... shows tracker video, so that's good. Video. Which had an accurate... position versus time graph. Showing constant velocity. Um... however, did not discuss main physics idea... which is... [inaudible].
S [18:41]: [laughs] Okay. So, the other what if question... so he addressed the second what if question... and answered correctly that the forces of... [inaudible] individual forces... cannot be determined, however, his explanation was incorrect.
S [19:22]: Invisible and tough to determine. Especially since if something is tough to determine, that means it can be determined so he just counter-argued himself. Um...When the correct answer, if I do remember, would be to explain forces such as friction um... weight force, et cetera, that we can measure.
S [20:12]: He also would have just [inaudible]
S [21:13]: Okay. He has just explained the most important lines of code. So if we go back to content models, he shows the code, and points out the important lines. No wall of code. So it's good.
S [21:41]: He explains... the significant lines, but... and... which variables... are included in each equation. But... still... has not told us what Newton's second law is... or why it is important in this experiment. Or why fnet is equal to zero.
S [23:25]: Okay... okay we're getting into content prediction discussion.
S [23:52]: Okay. Graphed... tracker data and... vpython... data. On same axes. That's good.
S [24:40]: Okay. Okay. [inaudible] Good explanation of... discrepancies between data...
S [25:34]: Whoop, there was no conclusion. Oh well. Production delivery, there were no problems. Spoke clearly... [inaudible]
S [25:57]: Okay. Um, do you want me to give ratings?
I [26:01]: Yes, if you could please assign one of the poor to excellent ratings for each item.
S [26:06]: Okay, so, I mean for production delivery, um... it's very good. Very good. There was nothing wrong with it, I could understand everything he was saying. Lemme go back to introduction—or organization structure. Um... I'd say it was good, there was... I... wasn't lost in my... chose to, place certain parts of the video next to each other. However, his introduction, well, I guess I [inaudible] because his introduction was not... there was no introduction [laughs]. And that kind of threw the whole video off because there was no explanation of the physics principles.
S [27:09]: Content models. Um... he did explain... each model. But there was no discussion of physics principles. So... that would put him down to... oh, well, he did say there was constant velocity. He didn't explain—he never said what Newton's second law was. So, I'd have to say poor
S [27:46]: Content prediction discussion. Hmm.
S [28:05]: Um... well, he did explain the discrepancies between data, and he showed all the graphs. Um... so I'm gonna say good. Even though... he had some problems. Okay.
S [28:28]: Content overall... he had some issues with the axes thing... at first... [muttering] Oh he had some problems with "invisible" and "tough to determine". Um... did address [inaudible]... I'm gonna say fair, because he got the first question right, not the second. Okay. So I'm done grading.
I [29:14]: Alright, excellent! So I'm going to send you the expert evaluation of that same video, uh, and I'd like you to just take a few moments to compare that to your evaluation, and react.
S [29:25]: Okay.
I [29:26]: So I've sent it.
S [29:38]: Introduction, organization structure. Fair. Oh. Wow. Poor. [inaudible] Good. Content model... fair. [mumbling]
S [30:05]: Um... okay. I don't agree with this one.
I [30:09]: Mm-hmm.
S [30:10]: But, that's... fine.
I [30:14]: Could you say that last part again? What you just said?
S [30:18]: I said I don't agree with the content overall rating. I... feel like we had... see, we had the exact same reasons... for giving him different grades, which makes the system not... good. [laughs] Because if you have an inadequate explanation for one question, how can you still get good? So... that's where I am.
S [30:50]: But other than that, I'm surprised. I was very similar to these grades.
I [30:57]: So—why, why is that—why is that surprising?
S [31:02]: Because like I said, I could have the same explanations as the grader, or as another student, and... I would give them a different um, grade from fair to excellent.
S [31:26]: And, or, um... for example, during the year, I would have my lab graded by three—or what was it, five?—students or something. All of them would give me—they would have the same comments, and would give me different grades. So I would have my TA look at it—completely different. But everybody had the same comments, so... [laughs] that's why I'm surprised that I got [close?], I had more similar gradings as the um...
I [32:00]: Can you tell me about, um, how your own calibration ratings—or sorry, calibration grades changed throughout the year?
S [32:11]: Um... I don't—wait, was the calibraiton one like the grade that I got for grading everyone else?
I [32:21]: Yeah so it was the grade that you got from how your ratings compared to the expert ratings for the calibratoin videos, and it was intended as a measure of how close your ratings were to the expert ratings.
S [32:34]: Um, I don't know. I didn't really look at them. But I felt like they stayed the same. They were kind of inconsistent throughout... like, uh... throughout the year.
S [32:52]: So I'd like to talk for alittle bit about, um... correct me if I'm wrong, but you're expressing some frustration at the unreliability of the ratings given the same comments. And this is something I'd like to explore. So you said that people were—the graders that you got were relatively consistent in giving the same comments, or at least you had the impression—
S [33:17]: Yeah, so everyone I felt like could tell what, what was good about a video, and what was not as good about a video.
I [33:26]: Mm-hmm.
S [33:27]: And so, some people I would say are more harsh than others. When... you know, if they saw one problem, somebody would... [give?] a fair when somebody else would give them a good.
I [33:44]: Do you think everyone was, um... I'd like to talk about people's, uh, ability to identify problems and identify good parts of videos, um, particularly with peers compared to experts. So do you think there's a difference in ability to pay attention to videos, or recognize things between the students and the instructors?
S [34:10]: Um... I mean I didn't really get a chance to... see like instructor's grading, except for the expert gradings.
I [34:22]: Well, those are the ones I'm talking about.
S [34:25]: Okay. So um... I would say... I mean, I guess it's kinda difficult to say because the instructor was, it was me comparing mine to what they wrote on one video, whereas... when I thought other students, when it would be five students... all grading my video. So like I couldn't see an instructor, um, instructor grade versus another students' grade on the same video. But I would say I was usually not consistent with what the um, what the expert [air quotes] grade was.
I [35:10]: Right.
S [35:11]: Yeah, and then... I mean... the—I have like, the students versus my TA, like my TA would give me... nineties or eighty-fives or whatever, and I would always get like, sixty-fives from my peers. So I felt like my—or like the students, um, were maybe more nitpicky than the instructors.
I [35:39]: Uh, let's talk about that for a bit. Um... by nitpicky, do you—let's just explore that for a bit. What do you mean by nitpicky, in particular?
S [35:52]: Um, maybe they were more harsh graders, or because they had just spent two weeks making their video, they were hoping that everyone else would spend—or, would [glitch] [spend?] just as much effort [in grading?] their video. Sss... uh, yeah I would just say that the students were—or they were just trained too well to look for um... these certain physics principles. It's also like, when you're grading five videos in a row, [glitch] poor, and you know immediately, this video doesn't have this, but the other video explain this better. So... whether you want to or not, you're comparing these five videos, not just the one video.
I [36:41]: And did you find yourself doing that?
S [36:45]: Um... I tried not to. But I would say that yeah, I probably did. I mean by the—the—if you hadn't watched any videos, and were grading the first one, you subconsciously give it an average score, and then the last video you... [glitch] first four. So it would probably be graded more harshly than the first video.
I [37:16]: The more... the more you grade—sorry, within a single grading session, you feel like you might be more inclined to give the later videos in that session a harsher grade?
S [37:27]: Um... yeah. I guess, depending on... what... what the quality of your five videos were.
I [37:34]: Mm-hmm.
S [37:36]: And yeah, I'd say probably.
I [37:37]: So you mentioned that you think, um, the ratings that other students give might be motivated in part by how much effort they feel they put into producing their own videos. Uh, can you talk a little bit about—a little bit more about that, and, uh in particular, how much effort you felt you put into your videos?
S [37:58]: So... um... when you grade your first video, all you have to compare it to were the calibration videos. And um... you're um... so... and, it even asks you, is this video better or worse than yours. [laughs] so I think the teachers kind of motivate you to compare them. And it asks you how many hours you put into it, so if somebody is—if you've seen that your video might have had a video or explained a concept better than another video, then "well they didn't have it and I did it" they—or "I didn't have this, they explained this better than me, I have to do that next time". And I knew somebody who—I think every week I had to grade one of my own videos.
I [38:58]: Mm-hmm.
S [38:59]: So I was comparing myself to the other four, to see if I had put as much effort as everyone else had.
I [39:09]: And—and did you?
S [39:11]: Um, yeah. I had. It was... there was also a problem with my video—like my microphone doesn't work when I'm screensharing or something, so I would always have everything written out on a... on a powerpoint. So it... was so, my videos would always explain them a little bit differently, and you could—when you're listening to someone's

voice, you cand kinda gauge how well they understand the material. So that was a little hard for me to compare. I think I did put in a good amount of effort on... my labs.
I [39:56]: Wonderful, so let's um... I want to take the conversation to, uh, something I noticed you doing at this video, when you were evaluating it. You were, um, you tended to watch for a few moments, pause it, type up some commentary, uh and then go back to the video and keep watching. Was this, uh, typical of how you would evaluate actual videos throughout the semester?
S [40:19]: Yeah. Because um, every... everything they say is important. Because I mean, you're on a time constraint, so you have to fit everything into your five minutes. And so, I want to make sure I remember exactly what they say. Um... and also, the videos are kind of organized according to the rubric. So it's easy to do a section before you forget about it at the end of the video.
I [40:50]: And had you always done this, uh done it this way. All the way since lab one?
S [40:56]: I don't know if I did it during lab one. But, after lab one, I found it to be the easiest way to evaluate them.
I [41:06]: And speaking of how you were doing things in lab one, um, so now you've had a full semester of experience making and evaluating these videos. You've done several, at least twenty videos—or rather sixteen practice videos at least, since the very beginning, to say nothing of the student videos. And the video that you evaluated right now is actually the same video you evaluated in the first interview session that we had. So, I'd like you to just, tell me a little bit about how you feel now as compared at the beginning of the semester, what do you think about these videos now, compared to how you did at the beginning of the semester.
S [41:48]: Um... like, how... what I think about them in terms of how I graded them?
I [41:53]: Yeah, how you grade them, what you pay attention to, what sort of things you want to see now as opposed to earlier in the semester.
S [42:02]: I mean, earlier in the semester, I don't know how many lab videos I had done, maybe one, um, but once you watch everyone else's, and then you make your own, you know how they're generally organized and how... [glitch] you [learn?] about the concepts that you're supposed to be portraying.
I [42:25]: I'm sorry you cut out for a second, could you go back about ten seconds?
S [42:31]: Yeah so after I had done... created more of my own videos, and watched other people's, it's easier to know how videos should be organized in order for somebody to understand the—the concepts that you're supposed to be teaching. So... I guess I can—I wouldn't say my, um, grading has changed as much as like, how quickly I can evaluate it. Um, and I know more of what to look for. When, like, if... if somebody doesn't state a physics principle at the beginning, it's hard to catch up at the end, [personally, I think?] as a viewer.
I [43:25]: Alright, good. Um... okay, we have some more topics for discussion which we might circle back to towards the end of the interview, but for now I'd like to move onto the next phase, which is watching a bit of observational data, um, and then discussing it. What I'd like you to do is to watch the ball rolling example, uh, video file that I sent you. Um, it's the same one we did at the beginning of the semester. It's a video of a ball rolling, and then the latter half of the video has tracker data. And I'd like you to just explain to me the physics of what's going on, and in particular, why the Tracker plot has the shape that it does.
S [44:12]: [muttering]
I [44:15]: And if you could please screenshare with me while you're doing that.
S [44:49]: [inaudible]
S [45:16]: Well. It looks like they're trying to get this ball rolling at a constant velocity. It... they have a... position versus time graph over here. Um... and [it's?] a little bit curved, only slightly out here—my roomate's here, let me adjust—I might have to migrate to my room.
I [45:41]: That's alright, you can take a moment to do that.
S [46:03]: [unintelligible conversation with person off-screen]
S [46:24]: Okay. So it looks like they were trying to do... a ball rolling at constant velocity. Um... they did a pretty good job because the, um, the [inaudible] almost straight slope. Um, and it could be bent because of like, frictional forces, and because they can't—or they didn't, um... measure the applied force on... the ball, with the hand pushing the ball. Um...
I [46:58]: Can you tell me a bit more about the force of the hand pushing the ball, and how it affects the shape of that position versus time curve?
S [47:10]: The force of the hand pushing the ball?
I [47:14]: You mentioned that they might not have measured the force of the hand pushing the ball, and I want to explore that a bit.
S [47:21]: I mean, when you just push a ball, you don't know how many... newtons you're applying, and you don't know the force of friction that the ground is applying. So... and like... what spots on the ground would be... would cause, you know, more friction, so you can't... predict exactly the... velocity would be constant, but like the average velocity... might be constant. [laughs]
I [47:56]: Uh, and what—what could the experimenter have done differently to make a line that's more straight?
S [48:03]: Um... well, you could... lemme just look back at the way they pushed the ball... [the ball's force?] is pretty consistent. Um...
S [48:27]: Well, to make a line that's more straight, you might choose a surface that... is... less rough. Or you might um... choose—like, they had their hand on the ball kind of for a while, so maybe she could use a force that is more quickly applied. But... I mean, for an experiment like this, we saw the data was relatively straight. Like, there wouldn't be a huge discrepancy between this data and a vpython model.
I [49:12]: Alright. Um, you mentioned... uh, just now, that the force was relatively consistent. The force of the hand pushing the ball was relatively consistent. What would have it have meant—what would the effect have been if there were an inconsistent push, and what do you mean by consistent and inconsistent?
S [49:36]: Well, if there were—if you keep your hand on the ball for more time, and you speed it up, it's—like your hand also has to move at a constant... rate for the ball to continue at a constant rate. So [if?] you speed your hand up, the ball will be more likely to accelerate.
I [50:02]: Uh, and how is that captured in the... sorry, how would that be captured in the shape of that plot there?
S [50:10]: Uh, the plot would not be a straight line. It would be curved. Well, like [laughs] it is. So.
I [50:18]: Um... let's see, so can you step me through uh... rewind the video a little bit until it just gets to the beginning of the tracker section. And talk to me about the force of the push from the hand in just like the first few frames of the video. How it affects the first few frames of the video.
S [50:43]: Okay.
I [50:44]: Yeah the first few data points there.
S [50:47]: I mean, they didn't start it until the ball... had gotten to this arbitrary origin that they chose. So, I mean, the force of the hand—the force of the hand is the readon it' moving, that's, it sets the speed that the ball is going to roll at. It—here, let me go back here to the [mumbles] data points. I mean... it... by just me looking at them, they look almost evenly spaced. So that means the force of the hand was... like... consistently applied over the time that it was on the ball. Um...
I [51:40]: And what is the force on the ball now?
S [51:43]: Right now? It's the frictional force that's going in the negative x direction, and the weight force on the ball, like the gravitation and the force of the Earth pushing up on the ball. The normal force.
I [52:03]: Um, can you tell me a little bit more about the role of friction in how this system evolves?
S [52:11]: The roll of friction?
I [52:13]: The sort of input that friction has, the r-o-l-e of friction.
S [52:17]: Oh, okay. The friction will slow the ball down, like there's—if there was no friction, then the ball would continue at the same speed as when it left the hand.
I [52:31]: Mm-hmm.
S [52:31]: So... friction is the reason that the ball has varying speeds.
I [52:39]: And what is the source of the frictional force that you're talking about?
S [52:44]: Um, the, I guess the unevenness of the surface that it's on. And the material that the ball's made out of. It looks like a tennis ball. Tennis balls aren't completely flat. So... that would affect it as well.
I [53:03]: And so if, um, if the materials are important, if the experimenter had chosen different materials, would the experimenter have been able to adjust the role of friction and the curve of the position versus time plot?
S [53:19]: Yes.
I [53:20]: How so?
S [53:21]: Like, if it—chose a, um, say a hockey table and a ping-pong ball.
I [53:34]: Mm-hmm.
S [53:35]: It might be more consistent because... you've um, lowered the force of friction.
I [53:43]: Mm-hmm.
S [53:45]: Cuz they're not as rough surfaces.
I [53:49]: Right. Um... and the... so let's talk a little bit about some of the more fundamental parts of the setup here. Um, the tracker is tracking the, a point that's located at the visual center of the ball there. Right.
S [54:12]: Yes.
I [54:13]: Tell me about that. Why... is the experimented justified in treating the entire ball as a single point located at its visual center?
S [54:25]: I mean, for this experiment, the center is the best place to pinpoint the ball, because it's the center of mass.
I [54:36]: Mm-hmm.
S [54:37]: And the ball is moving with its center of mass.
I [54:41]: The ball is moving with the center of mass, but the whole—not every part of the ball is moving at the same speed at the center of mass, right? Right, the...
S [54:54]: True...
I [54:54]: They're all—at any given moment in time, um, you know, the different points on the surface of the ball are moving at different directions at different speeds. So what is it that justifies the use of—or explain to me why we pick the center of mass when we're doing this sort of motion analysis.
S [55:15]: I guess the center of mass is like... an average almost, like that's the, the point at which... the, like all the other um... particles on the ball, like, I don't really know how to describe it, because I haven't had to do this in a while.
I [55:40]: Mm-hmm.
S [55:41]: [laughs] Like that's the point where all the other particles on the ball are centered around.
I [55:48]: Mm-hmm.
S [55:49]: And if you just picked an arbitrary point on the ball, it wouldn't be as... accurate to the whole, um... to the whole thing.
I [56:00]: What—what sort of data would you get if you picked, um, some other point on the ball?
S [56:07]: I mean, I guess it depends, but it would give you a... position, or it would give you a velocity, average velocity relative to that point. Not to the whole thing.
I [56:22]: Um, and finally, I'd like to ask a question about something you said during your video analysis. You were watching the lab video, and at a certain point, you laughed and said that, um... this was at the time when you were saying—you were typing, "he addressed the second what if question, and answered correctly that the individual forces cannot be determined, however his explanation was incorrect". And you explained [a bit about]
S [56:51]: [Oh yeah...]
I [56:51]: his reasoning there, he said "tough to determine"
S [56:53]: It was too tough to determine.
I [56:56]: And can you tell me a little bit more about that? Why did that make you laugh?
S [57:02]: Well first of all, he said that there were... forces—what was the question? Like, the wording of the question, like you had to determine the... um... can you determine the individual forces that make the net force zero?
I [57:22]: Uh yes, that's basically the wording.
S [57:26]: So he said no, you can't determine that, and he said no because it's too tough. Well if something's too tough, that means it—you actually can do it, it's just difficult. So... he counteracted himself. He said two different things. So that's why I laughed. But... he didn't say—he didn't actually say why, um, it was difficult. He just said that it was.
I [58:00]: Uh, and so how does that apply to this circumstance? Because you did say—you did enumerate a few forces, uh, you enumerated the force of the push of the hand, the force of gravity, the force of the Earth, and the force of friction on the ball. So how are you—why are you confident in naming those forces?

S [58:20]: Well... we know... based on our study in physics, that that's—that's what is affecting the ball, we've done like a million free body diagrams [laughs]. Enough to know that when somebody applies a force to that ball, the applied force—you can still see the effect of the applied force, but it's not being applied at that second. And so the reason that the ball isn't continuing like, Newton said, in constant motion is because there is another force going in the opposite direction, and that would be friction. And we know that the ball does have weight force going down. The reason it's not going into the ground is because there's another force uh going in the positive y direction by the Earth. So, that's how we know.
I [59:21]: Alright, and how do all of these forces in this case add up together? What is the ultimate result of all of these forces adding together?
S [59:31]: If this were an ideal situation, the net force would be zero.
I [59:35]: Mm-hmm. But it isn't an ideal situation. What is the net force?
S [59:38]: The net force would be the force of friction. Well... as it's moving in the positive x-direction, it's the force of, um, the applied force. But once it stops, that's when the force of friction has overcome the applied force.
I [59:57]: Okay, thank you very much.
[logistics etc.]

REFERENCES

[1] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, "Comparing the force and motion conceptual evaluation and the force concept inventory", *Physical review special topics - physics education research*, vol. 5, no. 1, pp. 010 105–8, Mar. 2009.

[2] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory", *The physics teacher*, 1992.

[3] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, "Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment", *Physical review special topics - physics education research*, vol. 2, no. 1, pp. 141–7, Mar. 2006.

[4] V. K. Otero and D. B. Harlowe, "Getting started in qualitative physics education research", 2009.

[5] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses", *American journal of physics*, 1998.

[6] M. D. Caballero, M. A. Kohlmyer, and M. F. Schatz, "Implementing and assessing computational modeling in introductory mechanics", *Physical review special topics - physics education research*, vol. 8, no. 2, p. 020 106, Aug. 2012.

[7] L. Ding, "Getting Started with Quantitative Methods in Physics Education Research", pp. 1–42, Jan. 2013.

[8] *Number of Students Enrolled in Physics and Physical Science Courses at TYCs and Institutions that offer a Bachelor's in Physics*, American Institute of Physics, Apr. 2013.

[9] *Bachelor's in Physics and STEM over Time*, American Physical Society, Apr. 2015.

[10] B. Schulz, "The importance of soft skills: Education beyond academic knowledge", *Journal of language and communication*, pp. 1–10, Sep. 2008.

[11] E. F. Redish, "Introducing students to the culture of physics: Explicating elements of the hidden curriculum", *Arxiv.org*, pp. 49–52, Aug. 2010. arXiv: `1008.0578v1 [physics.ed-ph]`.

[12] L. C. McDermott, P. S. Shaffer, and M. L. Rosenquist, *Physics by inquiry*. Wiley New York, 1996, vol. 1.

[13] R. Chabay and B. Sherwood, *Matter and Interactions 3rd Ed.* John Wiley and Sons, Inc., Jun. 2010.

[14] M. D. Caballero, "Evaluating and Extending a Novel Course Reform of Introductory Mechanics", PhD thesis, Georgia Institute of Technology, 2011.

[15] W. G. Chase, H. S. C. psychology, and 1973, "Perception in chess", *Elsevier*, vol. 4, no. 1, pp. 55–81, Jan. 1973.

[16] M. Chi, P. J. Feltovich, and R. G. C. science, "Categorization and representation of physics problems by experts and novices", *Wiley online library*, 1981.

[17] M. T. H. Chi, R. Glaser, and E. Rees, "Expertise in Problem Solving.", May 1982.

[18] J. Sweller, P. L. Ayres, S. Kalyuga, and P. Chandler, "The expertise reversal effect", *Educational psychologist*, vol. 38, no. 1, pp. 23–31, 2003.

[19] N. Falchikov and J. Goldfinch, "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks", *Review of educational research*, vol. 70, no. 3, pp. 287–322, Sep. 2000.

[20] M. van Zundert, D. Sluijsmans, and J. van MerriEnboer, "Effective peer assessment processes: Research findings and future directions", *Learning and instruction*, vol. 20, no. 4, pp. 270–279, Aug. 2010.

[21] *Edx*, `http://www.edx.org`, Accessed: 26-04-2018.

[22] *Coursera*, `http://www.coursera.org`, Accessed: 26-04-2018.

[23] M. Freeman and J. McKenzie, "Aligning Peer Assessment with Peer Learning in Large Classes: The Case for an Online Self and Peer Assessment System", in *Peer learning in higher education*, Jan. 2001, pp. 156–169.

[24] K. Topping, "Peer Assessment Between Students in Colleges and Universities", *Review of educational research*, vol. 68, no. 3, pp. 249–276, Sep. 1998.

[25] P. M. Sadler and E. Good, "The impact of self-and peer-grading on student learning", *Educational assessment*, vol. 11, no. 1, pp. 1–31, 2006.

[26] S. Murthy, "Peer-assessment of Homework Using Rubrics", *Aip conference proceedings*, vol. 951, no. 1, pp. 156–159, Nov. 2007.

[27] D. Brown, *Tracker video analysis and modeling tool*, http://web.archive.org/web/20160411171531/http://physlets.org/tracker/, Accessed: 2016-08-22, 2016.

[28] *Open Source Physics @ comPADRE*, http://web.archive.org/web/20160405010027/http://www.compadre.org/OSP/, Accessed: 2016-08-22, 2016.

[29] B. Sherwood, *VPython*, http://web.archive.org/web/20160710130011/http://vpython.org/, Accessed: 2016-08-23, 2016.

[30] M. R. Benioff and E. D. Lazowska, *Computational science: ensuring America's competitiveness*, 2005.

[31] J. Bergmann and A. Sams, *Flip your classroom*. International Society for Technology in Education, 2012.

[32] *Your World is Your Lab*, https://www.youtube.com/c/GTPHYSMOOC, Accessed: 2017-03-02.

[33] *Google Hangouts*, https://hangouts.google.com, Accessed: 2017-03-07.

[34] R. D. Luce and P. Suppes, "Representational measurement theory", *Stevens' handbook of experimental psychology*, 2002.

[35] S. S. Stevens, "Handbook of experimental psychology.", 1966.

[36] G. Casella and R. Berger, *Statistical inference*, ser. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002, ISBN: 9780534243128.

[37] T. Swinscow and M. J. Campbell, *Statistics at square one*, 2002.

[38] K. Kelley and K. J. Preacher, "On effect size.", *Psychological methods*, vol. 17, no. 2, pp. 137–152, 2012.

[39] C. J. Ferguson, "An effect size primer: A guide for clinicians and researchers.", *Professional psychology: Research and practice*, vol. 40, no. 5, pp. 532–538, 2009.

[40] J. L. Rodgers, W. N. T. A. Statistician, and 1988, "Thirteen ways to look at the correlation coefficient", *Taylor & francis*, vol. 42, no. 1, pp. 59–66, Feb. 1988.

[41] D. A. Darling, "The Kolmogorov-Smirnov, Cramer-von Mises Tests", *The annals of mathematical statistics*, vol. 28, no. 4, pp. 823–838, Dec. 1957.

[42] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. Wiley, 1999.

[43] R. C. S. Journal and 2012, "What hypotheses do" nonparametric" two-group tests actually test?", *Econ.uzh.ch*,

[44] A. Agresti, *An Introduction to Categorical Data Analysis, 1996*, 1996.

[45] *SciPy: Mann-Whitney U*, `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html`, Accessed: 11-04-2018.

[46] *SciPy: Chi-Squared Test with Contingency Table*, `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html`, Accessed: 11-04-2018.

[47] L. Bouzidi and A. Jaillet, "Can online peer assessment be trusted?", *Journal of educational technology & ...*, 2009.

[48] C. Piech, J. Huang, Z. Chen, C. Do, A. Ng, and D. Koller, "Tuned models of peer assessment in MOOCs", *Arxiv preprint arxiv:1307.2579*, 2013.

[49] E. Price, F. Goldberg, S. Patterson, and P. Heft, "Supporting scientific writing and evaluation in a conceptual physics course with Calibrated Peer Review", vol. 1513, pp. 318–321, 2013.

[50] R. Robinson, "Calibrated Peer Review™: An Application To Increase Student Reading & Writing Skills", *The american biology teacher*, vol. 63, no. 7, pp. 474–480, 2001.

[51] *BayesRank*, New York, New York, USA: ACM Press, 2015.

[52] D. M. A. Sluijsmans, S. Brand-Gruwel, J. J. G. van Merriënboer, and R. L. Martens, "Training teachers in peer-assessment skills: effects on performance and perceptions", *Innovations in education and teaching international*, vol. 41, no. 1, pp. 59–78, Jun. 2007.

[53] H. Smith, A. Cooper, and L. Lancaster, "Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development", *Innovations in education and teaching international*, vol. 39, no. 1, pp. 71–81, Jan. 2002.

[54] W. Cheng and M. Warren, "Making a Difference: Using peers to assess individual students' contributions to a group project", *Teaching in higher education*, vol. 5, no. 2, pp. 243–255, Aug. 2010.

[55] S.-Y. Lin, S. S. Douglas, J. M. Aiken, C.-L. Liu, E. F. Greco, B. D. Thoms, M. D. Caballero, and M. F. Schatz, "Peer Evaluation of Video Lab Reports in an Introductory Physics MOOC", in *2014 physics education research conference*, American Association of Physics Teachers, Apr. 2015, pp. 163–166.

[56] H. M. Cooper, *Synthesizing research: A guide for literature reviews*. Sage, 1998, vol. 2.

[57] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature", *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[58] *Python programming language*, `https://www.python.org`, Accessed: 11-04-2018.

[59] *Webassign*, `http://www.webassign.com`, Accessed: 11-04-2018.

[60] *Perl programming language*, `https://www.perl.org`, Accessed: 11-04-2018.

[61] *SWAPR on github*, `https://github.com/GatechVIP/gatech-swapr-server-node`, Accessed: 11-04-2018.

[62] *Georgia Tech Vertically Integrated Projects*, `http://www.vip.gatech.edu`, Accessed: 11-04-2018.

[63] *node.js programming environment*, `https://nodejs.org/en/`, Accessed: 11-04-2018.

[64] *PostgreSQL: The World's Most Advanced Open-Source Database*, `https://www.postgresql.org`, Accessed: 11-04-2018.

[65] *AngularJS*, `https://angularjs.org`, Accessed: 11-04-2018.

[66] H. Luo, A. C. Robinson, and J. Park, "Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects.", *Journal of asynchronous learning*, 2014.

[67] S. S. Douglas, S.-Y. Lin, J. M. Aiken, B. D. Thoms, E. F. Greco, M. D. Caballero, and M. F. Schatz, "Peer Evaluation of Video Lab Reports in a Blended Introductory Physics Course", in *2014 physics education research conference*, American Association of Physics Teachers, Apr. 2015, pp. 75–78.

[68] S. S. Douglas, J. M. Aiken, S.-Y. Lin, E. F. Greco, E. Alicea-Muñoz, and M. F. Schatz, "Peer assessment of student-produced mechanics lab report videos", *Physical review physics education research*, vol. 13, no. 2, pp. 257–14, Nov. 2017.

[69] R. K. Thornton and D. R. Sokoloff, "Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula", *American journal of physics*, vol. 66, no. 4, pp. 338–352, Apr. 1998.

[70] K. A. Ericsson and H. A. Simon, “Verbal Reports as Data”, *Psychological review*, vol. 87, no. 3, pp. 215–251, 1980.

[71] T. Claire, T. L. Tippett, and A. Boudreaux, “How accurate are physics students in evaluating changes in their understanding?”, in *2015 physics education research conference*, American Association of Physics Teachers, Dec. 2015, pp. 87–90.

[72] J. C. A. J. o. physics and 1982, “Students’ preconceptions in introductory mechanics”, *Physics.utoronto.ca*,

[73] D. Hammer, A. Elby, R. S. T. o. l. from, and 2005, “Resources, framing, and transfer”, *Books.google.com*,

[74] P. A. Carlson and F. C. Berry, “Calibrated Peer Review and assessing learning outcomes”, *Frontiers in education conference*, 2003.

[75] E. R. J. o. Algorithms and 1990, “NP-complete stable matching problems”, *Elsevier*, vol. 11, no. 2, pp. 285–304, Jun. 1990.

[76] Z. Fang, “Scientific literacy: A systemic functional linguistics perspective”, *Science education*, vol. 89, no. 2, pp. 335–347, 2005.

[77] B. J. Guzzetti and W. O. Williams, “Gender, text, and discussion: Examining intellectual safety in the science classroom”, *Journal of research in science teaching*, vol. 33, no. 1, pp. 5–20, Jan. 1996.

[78] R. L. L. i. society and 1973, “Language and woman’s place”, *Cambridge.org*,

[79] S.-Y. Lin, J. M. Aiken, D. T. Seaton, S. S. Douglas, E. F. Greco, B. D. Thoms, and M. F. Schatz, “Exploring University Students’ Engagement with Online Video Lectures in a Blended Introductory Mechanics Course”, *Arxiv.org*, Mar. 2016. arXiv: `1603.03348v1 [physics.ed-ph]`.

[80] *Piazza*, `https://piazza.com/signup`, Accessed: 11-04-2018.

VITA

Scott Douglas was born on July 5, 1987, in Atlanta, Georgia, to Bruce and Leigh Douglas. His education in Atlanta public schools led him to pursue a B.A. in physics at Colgate University in upstate New York, and the cold Yankee winters led him to pursue a Ph.D. in physics back down in Atlanta at Georgia Tech. He began his graduate career as a physicist in Dr. Steve Harvey's Computational Structural Biology lab, and concluded it as a researcher in Dr. Mike Schatz's Physics Education Research group.

While working in the Schatz PER group, Scott animated a series of physics lecture videos, led a team to develop peer assessment software, and spent many hours elaborating ideas about student scientific communication and physics understanding.

Scott married Anat Fintzi on September 6, 2014, while Anat was earning her M.S. in cognitive neuroscience at Vanderbilt University and Scott was studying at Georgia Tech. With seven years of long-distance relationship behind them, they now live together with their two cats in their home in Decatur.