BEYOND AGGREGATED DATA: A STUDY OF GROUP DIFFERENCES IN CONCEPTUAL UNDERSTANDING AND RESOURCE USAGE IN AN UNDERGRADUATE DYNAMICS COURSE

by

Nick Alan Stites

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Engineering Education

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Edward Berger, Chair

School of Engineering Education

Dr. Jennifer DeBoer

School of Engineering Education

Dr. Allison Godwin

School of Engineering Education

Dr. Jeffrey F. Rhoads

School of Mechanical Engineering

Approved by:

Dr. Donna Riley

Head of the Graduate Program

To my family.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my wife, Emily, for all of her love and support. I would have never been able to pursue or complete this degree without you. This has been a long journey, but I loved travelling it with you. I would also like to thank my children, Addison and Logan. Your smiles and laughs constantly remind me to embrace the moment and to be thankful for those around me. To my parents and sisters—thank you. You have been instrumental in shaping who I am today. My entire family has always believed in me and supported me, and this PhD experience has been no different.

I would also like to thank my advisor, Dr. Edward Berger. Thank you for caring about me professionally and personally. You understood the challenges of being a student, researcher, father, and husband. You gave me the freedom to forge my own path, and you were always encouraging, calm, and insightful. I am also grateful to my committee members: Dr. Jennifer DeBoer, Dr. Allison Godwin, and Dr. Jeffrey F. Rhoads. Thank you for your expertise and continually challenging me to be a better researcher.

The second chapter of this dissertation would not have been possible without the contributions of Dr. Kerri Douglas. Thank you for providing the expertise to organize and frame my validation study to tell a coherent story. I would also like to thank Dr. Alice Pawley for helping me explore and contribute to the field of gender studies in engineering education. You have been extremely understanding and generous with your time. Finally, thank you to Dr. Amy C. Moors, Dr. Cheryl Cooky, and Dr. Sharra Vostral for their input and expertise on the gender-bias analysis in Chapter 2. I hope all of you are as proud of the chapter as I am.

To Dr. Charles M. Krousgrill, Jr.—thank you. You were always willing to help with my research, and you provided me with an incredible mentored-teaching experience. My semester of teaching with you was one of the best experiences of my PhD. Thank you for being such a great mentor and friend.

I would also like to thank the entire Freeform research team. Many aspects of this dissertation, from data collection to analysis to writing, reflect a team effort. Angela Goldenstein, thank you for your positive attitude and leadership. Ruth Rothstein, thank you for your project support and always being willing to call me into team meetings. Rohit Kandakatla, thank you for developing the coding scheme for the student interviews, and Garrett Stocksdale, thank you for your tireless coding efforts. David Evenhouse, thank you for your contributions on the aDCI analysis. Craig Zywicki, thank you for maintaining such an extensive data set and for always accommodating my questions and requests. Nimit Patel, Claudio Freitas, and Austin Zadoks, thank you for helping collect data and for reviewing my analyses.

Lastly, I would like to acknowledge that the College of Engineering provided funding for my research assistantship and that this dissertation is based on work supported by the National Science Foundation under Grant No. DUE-1525671. Any opinions, findings, and conclusions or recommendations expressed in this material are my own and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

LIST OF TABLES x
LIST OF FIGURES xii
ABSTRACT xiv
CHAPTER 1. DISSERTATION OVERVIEW 1
- Introduction 1
  - What is Freeform? 1
  - Purpose of Dissertation 4
- Background 5
  - Relating This Research to Other Freeform Research 5
  - Why Study Dynamics? 5
- Overall Research Approach 6
  - Research Methods 6
  - Research Setting and Context 7
  - Participants 8
  - Researcher Positionality 9
- Chapter Organization 10
  - Chapter 2. A Validation and DIF Study of the aDCI 10
  - Chapter 3. Identifying and Understanding Resource-Usage Patterns 11
  - Chapter 4. Relating Resource-Usage Patterns to Academic Achievement 11
- Contribution 12

CHAPTER 2. A VALIDATION AND DIF STUDY OF THE ADCI 14
- Abstract 14
- Introduction 15
  - Purpose of the Study and Research Questions 17
- Literature review 19
  - Reliability, Validity, and Fairness 19
  - Sources of Gender Bias in CIs 20
- Background 22
  - aDCI Development 23

Methods 24
Participants and Data Collection 24
Data Analyses 26
Results 30
Descriptive and Correlation Statistics 30
Hypothesis 1: A Single-Factor Latent Structure 33
Hypothesis 2: Items of Appropriate Difficulty and Discrimination 34
Hypothesis 3: Correlation with Similar Measures of Conceptual Understanding 37
Hypothesis 4: Measurement Invariance Across Genders 38
Gender Bias in the aDCI 42
Description of the Biased Items 42
Bias in Q3 42
Bias in Q6 43
Discussion 44
Review of Purpose and Results 44
Fairness Implications 46
Limitations and Future Work 46
Conclusion 47
Acknowledgements 48
CHAPTER 3. A CLUSTER-BASED APPROACH TO IDENTIFYING AND UNDERSTANDING STUDENTS’ ARCHETYPICAL RESOURCE-USAGE PATTERNS 50
Abstract 50
Introduction 51
Purpose of Study 53
Study Context 54
Conceptual Framework 55
Importance of Help-Seeking Behaviors 55
Help Seeking and Self-Regulated Learning 55
Expectancy-Value Model for Help Source Selection 56
Prior Dynamics Results and the EV Framework 57
Methods 58

Participants 58
Data Sources 59
Data Analysis 61
Results 63
Cluster Analysis of Survey Data 63
Thematic Analysis of Interview Data 66
HSB Orientations of Each Cluster 76
Discussion 80
RQ1: Patterns of Resource Usage 80
RQ2: Understanding How and Why 81
Implications for Practice 82
Implications for Research 83
Limitations and Future Work 84
Conclusion 85
Acknowledgements 86
CHAPTER 4. RELATING RESOURCE-USAGE PATTERNS TO ACADEMIC ACHIEVEMENT 87
Abstract 87
Introduction 88
Purpose of Study 89
Conceptual Framework 91
General Model of Learning 91
Resource Usage and Achievement 92
Background 94
Dynamics Resources 94
Resource-Usage Patterns 96
Methods 97
Participants 97
Data Sources 99
Data Analysis 103
Results 104

Descriptive and Correlation Statistics 104
Group Comparisons 104
Multiple Regression Analysis 111
Discussion 114
Review of Purpose and Results 114
Resource-Usage Behaviors of C_8 115
Resource-Usage Behaviors of C_7 116
Suggestions from Other Clusters 116
Implications for Practice 117
Implications for Research 118
Limitations and Future Work 118
Conclusion 119
Acknowledgements 120
CHAPTER 5. CONCLUSION 122
Review of Purpose and Results 122
Contribution 124
Contribution 1: An-Argument-Based Validation Approach for CIs 124
Contribution 2: Awareness and Template for Evaluating the Fairness of CIs 125
Contribution 3: Holistic, Cluster-Based Approach to Studying HSBs 125
Implications 125
Limitations and Future Work 126
REFERENCES 128
APPENDIX A. FULL-STUDY CONSENT FORM 143
APPENDIX B. INTERVIEW CONSENT FORM 145
APPENDIX C. RESOURCE USAGE SURVEY 148
APPENDIX D. MOTIVATION AND TASK VALUE SURVEY 164
APPENDIX E. IJEE COPYRIGHT PERMISSION 167
APPENDIX F. LIST OF ACRONYMS USED IN CHAPTER 2 168
APPENDIX G. ADDITIONAL METHOD DETAILS FOR CHAPTER 3 169

LIST OF TABLES

Table 1. Demographics of students who enrolled in Dynamics from Spring 2015-Spring 2018 (*N* = 2037). 9
Table 2. An overview of the analytical methods used in this study and the purpose of each of the hypotheses. 18
Table 3. A gender-orientation framework for evaluating items on the aDCI for gender bias (Rennie & Parker, 1993). 22
Table 4. Description of the concepts assessed by each item of the aDCI (using verbatim descriptions from Gray et al. (2005)). 24
Table 5. Demographics of the sample (N = 1250). 26
Table 6. Low correlation coefficients (lower diagonal with standard errors in the upper diagonal) between items of the aDCI illustrated the broad nature of the concepts assessed by the aDCI. 31
Table 7. Goodness of fit and model comparison statistics for testing measurement invariance of the aDCI across men and women. 32
Table 8. Summary of the IRT parameters and the answer distributions for lower-ability students who answered three or fewer questions correctly on the aDCI. Correct answers are bolded. 36
Table 9. Strong correlations (coefficients in the lower diagonal with standard errors in the upper diagonal) between different measures of the students' conceptual understanding suggest that they all may be measuring the same construct. 38
Table 10. The correlations coefficients between the aDCI items for men (below the diagonal) and women (above the diagonal). 39
Table 11. The thresholds for nine of the eleven items were invariant across gender; Q3 and Q6 had unequal thresholds (bolded), which suggested that they were potentially biased against women. A lower threshold corresponds to a higher proportion of students in that group answering the item correctly. 41
Table 12. Demographic characteristics of the survey and interview participants. 59
Table 13. A description of the nine resources included on the end-of-semester survey and the median frequency with which students used the resource (N = 581). 60

Table 14. Examples of adaptive HSBs for each cluster. The bolded text indicates key phrases related to adaptive HSBs. 78

Table 15. Examples of adaptive HSBs for each resource listed on the survey. The bolded text indicates key phrases related to adaptive HSBs. 79

Table 16. A description of the nine resources included on the end-of-semester survey and the median frequency at which students used the resource (N = 581; adapted from Stites, Berger, et al., in review-a). 95

Table 17. The demographic characteristics of the samples used in this study. The sample used for the regression analysis is a subset of the clustering-analysis sample because of data availability 99

Table 18. Descriptive statistics and group comparisons for the achievement metrics. Clusters with means that statistically differed ($p < 0.050$) from that of the rest of the sample are bolded. All *p*-values were multiplied by nine as a Bonferroni adjustment. 106

Table 19. Descriptive statistics and group comparisons for ordinal and continuous control variables. Means that statistically differed ($p < 0.050$) from that of the rest of the sample are bolded. All *p*-values were multiplied by nine as a Bonferroni adjustment. 106

Table 20. Descriptive statistics and group comparison for the dichotomous control variables. Cluster proportions that statistically differed ($p < 0.050$) from that of the rest of the sample are bolded. The *p*-values were multiplied by nine as a Bonferroni adjustment. 107

Table 21. Correlation coefficients (lower triangle) and their respective p-values (upper triangle) for control and achievement variables. Statistically-significant ($p < 0.050$) relationships are bolded. 108

Table 22. The estimated regression coefficient for models comparing academic achievement across clusters. Estimates that were significant ($p < 0.050$) are bolded. 112

LIST OF FIGURES

Figure 1. The single-factor structural model fit the aggregated data from men and women well, as tested in Hypothesis 1. The numbers on the arrows from the latent construct (conceptual understanding of dynamics) to the items (Q1-Q12) represent factor loadings, which in this study are equivalent to correlation coefficients. The numbers below the items indicate the proportion of unexplained variance in each item. 33

Figure 2. A 3PL IRT model was used to fit each aDCI item to determine the discrimination, difficulty, and guessing parameters. The difficulty (inflection point of the curve and the latent ability level that bisects the sample) and discrimination (slope at the inflection point) are indicated for Q01, and the guessing parameter is illustrated for Q12. The random-guessing probability is 0.2 because all items have five answer choices. 35

Figure 3. An item-person map (or Wright map) of the ability scores of the participants and the difficulty values for the 12 items of the aDCI. A logit (vertical axis) is the natural log transformation of 1) the odds ratio for answering an item correctly, or 2) the ratio of a student's ability divided by one minus their ability (Wilson, 2005). An item for which a student, in general, would have a 50% chance of answering correctly would have a logit of zero, and an average-performing student with an ability of zero would have a logit of zero. 37

Figure 4. The students' total aDCI scores can be used as a measure of conceptual understanding because the aDCI scores are linearly related and highly correlated with the latent factor score from the CFA. 38

Figure 5. The test results of Hypothesis 4 showed the aDCI scores to have invariant form, factor loadings, and thresholds across men and women. All of the factor loadings were statistically significant with p-values < 0.001. 41

Figure 6. A self-regulated-learning perspective of HSBs underpins the conceptual framework of this study (Karabenick & Berger, 2013). 56

Figure 7. An expectancy-value model for help source selection (Makara & Karabenick, 2013).57

Figure 8. The average values for how frequently the students within a cluster used each of the nine resources. The sample sizes with the "int" subscripts indicate the number of interviewees in each cluster. 64

Figure 9. A model for motivation and self-regulated learning for college students. Adapted from Pintrich and Zusho (2007). 92
Figure 10. The Dynamics students exhibited nine archetypical resource-usage patterns (Stites, Berger, et al., in review-a). The “reg” subscript indicates the number of students in each cluster for the regression analysis, and the “clust” subscript designates the sample size for the cluster analysis. The two samples differ because of data availability........ 97
Figure 11. Cluster distributions of the three achievement metrics: overall course grade, performance on the problem-solving (PS) questions of the exams, and performance on the conceptual questions of exams. Clusters with central tendencies that differed ($p < 0.050$) from that of rest of the sample are bolded. 109
Figure 12. Cluster distributions for self-efficacy and cumulative GPA at the start of the semester in which students took Dynamics. Self-efficacy was measured on a 1-7 Likert scale that corresponded with increasing levels of self-efficacy. Cumulative GPA was measured on a 4.0 scale (A = 4.0). Clusters with central tendencies that differed ($p < 0.050$) from that of rest of the sample are bolded. 110

ABSTRACT

Author: Stites, Nick A., PhD
Institution: Purdue University
Degree Received: December 2018
Title: Beyond Aggregated Data: A Study of Group Differences in Conceptual Understanding and Resource Usage in an Undergraduate Dynamics Course
Committee Chair: Edward Berger

As pedagogical innovations continue to be developed and adopted in engineering education, it is important to understand how these innovations affect the students' experiences and achievements. A common data analysis practice when evaluating educational innovations is to aggregate the data from all of the students together. However, this data aggregation inherently biases the results toward the characteristics of the dominant student group, leaving the experiences of minority groups largely unexplored. In this dissertation, I investigate the students' experiences and achievements in an undergraduate dynamics course, and I intentionally use analysis methods that disaggregate the data to better understand the behaviors and performance of smaller subgroups of students, not just the majority.

This dissertation presents three studies that examine: 1) the validity, reliability, and fairness of a standardized set of conceptual questions on the final exam, with a focus on gender fairness, 2) how and why the students use the available resources, and 3) how the students' holistic resource usage patterns relate to their academic achievement. My motivation for choosing these studies was that conceptual assessments and customized resources are two key components of the learning environment for the dynamics course. However, the quality of the conceptual exam questions used for the course had yet to be evaluated. Similarly, the learning environment for the course incorporates many customized resources, including a custom-written "lecturebook" (a hybrid of a textbook and a workbook) and an extensive online library of videos, but little was known about how the students used these resources, or how the students' pattern of resource usage related to their performance in the course.

The first study in this dissertation used multiple-group confirmatory factor analysis to investigate item-level gender bias in a 12-item Abbreviated Dynamics Concept Inventory (aDCI), which was a set of standardized conceptual questions included on the final exam. The results suggested that two items were slightly biased against women, with stereotypically-

masculine contexts and content as possible sources of the bias. The bias in the aDCI items likely unfairly lowered some women's final exam scores, highlighting the need for engineering educators to consider the fairness of their assessments.

The second study used a cluster analysis of survey responses to identify nine archetypical patterns of resource usage, all of which differed from the average resource-usage pattern of the aggregated sample. An analysis of forty-four student interviews, organized by resource-usage cluster, determined that students exhibited their resource-usage behaviors largely because of how they perceived the resource's availability, accessibility, and quality. The results illustrate that there is no "typical" way in which the students used the resources, so it is important for instructors to consider a wide array of usage behaviors when designing a course's learning environment and resources.

The third study utilized a multiple regression analysis to find that *on average* a student's resource-usage pattern is not related to their achievement when controlling for many other demographic, cognitive, and non-cognitive factors that can affect resource usage and performance. However, two individual resource-usage patterns were significantly related to achievement. Students who primarily used their lecturebook and their peers for support performed better than their similar peers in other resource-usage clusters. Conversely, students who rarely used their lecturebook had lower course grades than their peers. Drawing from the results of the second study, general study-habit suggestions for the students in the course were extracted from the qualitative themes found in the interviews of the students in these two clusters.

Overall, the results of these three studies highlight how the experiences and achievements of smaller groups of students would go unnoticed if analytical methods that only utilized aggregated data were used. While the setting of this research is specific to the assessments and resources of a given dynamics course, the methods used to disaggregate the data to gain insights about different subgroups of students are applicable to many engineering education contexts. My hope is that this work inspires more researchers to consider the experiences of all students, not just those of the majority.

CHAPTER 1. DISSERTATION OVERVIEW

Introduction

As pedagogical innovations continue to be developed and implemented in engineering education, it is important to understand the impact that these innovations have on the students' experiences and learning. However, when new instructional methods or resources are evaluated, researchers often aggregate data from all the students together for analysis. These analyses of aggregated data inherently yield results that primarily reflect the characteristics of the majority and can overlook important differences in the data, such as the experiences of students in a smaller subgroup of students. For example, in engineering in the USA, women are outnumbered by men by about a factor of four (National Science Board, 2018), so the evaluation of a pedagogical innovation in engineering education is dominated by the experiences of men and offers very little information about the experiences of women. Researchers of design have argued for years that it is important to consider all users of a product, not just the average or stereotypical user (Cooper, Reimann, & Cronin, 2007). When this argument is applied to education, it highlights the need to understand how all of the students (the users) experience an educational activity (the product). Therefore, this dissertation uses analysis methods that disaggregate the students' data to better understand the experiences and achievements of smaller subgroups of students in an undergraduate engineering course that utilizes an active, blended, and collaborative learning environment called Freeform.

What is Freeform?

Active, Blended and Collaborative Learning Environment

In 2010, two professors in the School of Mechanical Engineering at Purdue University formally implemented a new pedagogical environment, called Freeform, in an undergraduate course on dynamics and vibrations, which I hereafter refer to as Dynamics, with a capital "D." Freeform incorporates aspects of active, blended, and collaborative learning to increase the students' learning and perceptions of the class. Active learning refers to times when the student is physically active (Chi, 2009; Freeman et al., 2014), blended learning combines in-class and online learning (Bernard, Borokhovski, Schmid, Tamim, & Abrami, 2014; Means, Toyama, Murphy, & Baki, 2013), and collaborative learning involves students working together to accomplish a shared goal (D. W. Johnson, Johnson, & Smith, 1991; K. A. Smith, Johnson, & Johnson, 1981). Components of the learning environment that facilitate and foster this type of learning include a custom-written “lecturebook” that is a combination of a textbook and a workbook, an extensive online video library that includes solution videos for all of the lecturebook example problems and every homework problem, and peer-to-peer collaboration in the classroom and online via a discussion forum where students can ask questions about their homework or the course material. The Freeform environment also leverages the availability of a tutorial (or “help”) room that is staffed by teaching assistants (TAs) and available over 40 hours per week.

Conceptual Understanding and Problem-Solving Skills

One of the educational philosophies that underpins the Freeform learning environment is the importance of both conceptual understanding and problem-solving skills, which can be considered a combination of procedural and conceptual knowledge (Bohle Carbonell, Stalmeijer, Könings, Segers, & van Merriënboer, 2014; Rittle-Johnson & Schneider, 2015). Conceptual knowledge is knowledge of “abstract and general principles”, and procedural knowledge is knowledge of “a series of steps, or actions, done to accomplish a goal” (Rittle-Johnson, Schneider, & Star, 2015, p. 588). Research has shown that procedural and conceptual knowledge influence a student’s ability to adapt and use existing knowledge in a new and unfamiliar application (Bohle Carbonell et al., 2014; Hatano & Inagaki, 1984). In many engineering science courses, conceptual knowledge is not emphasized, but in Dynamics, it is of such importance that every Dynamics exam includes problem-solving questions *and* conceptual questions.

Since Spring 2015, instructors and researchers incorporated an Abbreviated Dynamics Concept Inventory (aDCI) into the Dynamics final exam as a replacement for custom-written conceptual questions that changed every semester. The adoption of the aDCI enabled the comparison of the students’ conceptual understanding across sections and semesters. A concept inventory (CI) is a multiple-choice assessment that requires little to no calculations. It is intended to measure a student’s conceptual understanding of the material, rather than their ability to solve a complex mathematical problem. The aDCI is a 12-item subset of the 29-item Dynamics Concept Inventory (DCI; Gray et al., 2005) and is similar to popular physics CIs, like the Force Concept Inventory (FCI; Hestenes, Wells, & Swackhamer, 1992). Two versions of the aDCI are used in this research, as further explained in Chapter 4, and the first version includes two questions that were copied directly from the FCI. Prior research has identified gender biases in the FCI, and this close relationship between the aDCI and the FCI was part of my motivation for conducting a gender fairness study of the aDCI, as presented in Chapter 2.

Success and Adoption of Freeform

Since the formal inception of Freeform for Dynamics in Fall 2010, the students' overall performance in Dynamics has improved markedly, and the students' opinion of the course is quite positive (Rhoads, Nauman, Holloway, & Krousgrill, 2014). The rate at which students earn a D, F, or withdraw from Dynamics (the DFW rate) has decreased from ~20% in Spring 2010 to as low as 9% in Spring 2014 (DeBoer et al., 2016). With over 500 students completing dynamics annually, the reduction in DFW rate translates to more than 50 additional students passing Dynamics and progressing toward graduation—a known factor in a student's decision to persist in engineering (National Science Board, 2016, p. 43). The success of this course transformation has captured the attention of provosts, deans, and faculty at Purdue University and other domestic and international institutions. The Freeform environment is being used to teach undergraduate dynamics courses at Purdue University, Trine University, and McGill University (Canada), and a nascent version of Freeform is being used at Purdue University and Universidad del Norte (Colombia) to teach Mechanics of Materials.

Dynamics, with its increased student performance since implementing Freeform, serves as a model for transforming engineering classes that suffer from archaic pedagogies and poor student performance. However, prior to expanding Freeform into other courses or institutions, it is prudent to better understand the experiences and achievements of all of the students in Dynamics, the most mature instantiation of Freeform, rather than solely evaluating the implementation of Freeform by DFW rates or other statistics that stem from aggregated data.

Purpose of Dissertation

The overall goal of this dissertation is to illustrate the value and importance of using analysis methods that disaggregate the data so that researchers and instructors can better understand the unique experiences of smaller subsets of students. This dissertation focuses on two aspects of the students' experiences in Dynamics: 1) their performance on the aDCI, and 2) their holistic resource-usage behaviors and how those behaviors relate to their academic achievement in the course. I chose these two areas of focus because conceptual assessments and customized resources are two key components of Dynamics and the Freeform learning environment.

My motivation for investigating the students' performance on the aDCI is to determine if differences in aDCI scores across genders, as identified in previous research (Prebel, Stites, Berger, Rhoads, & DeBoer, 2017), could be reflective of bias in the aDCI against women. Prior validation studies identified gender biases in physics concept inventories, including the FCI, which by extension means that the aDCI could also be gender biased. I am also concerned that the semester-specific exam questions could be gender-biased because their format and content are similar to the aDCI and FCI, but the aDCI was the only set of conceptual (or problem-solving) exam questions that were used for multiple semesters. Therefore, the only exam questions for which I have enough student responses to conduct a validation study are those from the aDCI. Nonetheless, my expectation is that a validation and gender fairness study of the aDCI will expose general suggestions about how to create fair conceptual exam questions that can be applied to the development of semester-specific exam questions for Dynamics and other engineering courses.

The students' resource-usage behaviors interest me because so much effort has gone into the development of customized resources for Dynamics and the Freeform environment, but very little is known about how the students actually use the resources. What is known about the students' resource usage is specific to a given resource and stems from aggregated data. For example, from website analytics, it is known that students frequently watch the online solution videos for the example problems that are included in the lecturebook (Rhoads et al., 2014). Also, the general level of participation on the discussion forum can be inferred from the number of student posts, and the utilization of the tutorial room can be approximated by the TAs who staffed the tutorial room. Very little, however, is known about how the students use multiple resources in combination, or if certain types of students tend to use certain types of resources. Overall, my research on the students' resource-usage behaviors aims to clarify what resources students are using, how they are using them, and if their resource-usage pattern relates to their performance in the Dynamics.

Background

Relating This Research to Other Freeform Research

This dissertation is part of a larger research endeavor to better understand the experiences of students and instructors in the Freeform learning environment. The larger Freeform project has five research questions that ask about 1) how students' engagement patterns with the resources relate to their performance, 2) how background factors influence engagement and performance, 3) how students and faculty perceive the Freeform environment, 4) how Freeform fosters a sense of community, and 5) how group-level factors (e.g., instructor or institution) influence the students' performance and engagement. The aDCI validation study in Chapter 2 provides information on how a student's background factors influence their performance, and Chapters 3 and 4 investigate the students' resource-usage patterns and their relationship with performance.

The decision to limit the scope of this dissertation to the performance (based on course grades and assessment scores) and resource usage of students in Dynamics risks overlooking important experiential differences in alternative measures of engagement or academic success. The affective, attitudinal, and emotional dimensions of the students' experiences in Dynamics are studied as part of the larger Freeform project, but they are not the focus of this work.

Why Study Dynamics?

Dynamics is a sophomore-level, challenging, gateway course for many engineering majors. To be successful in Dynamics, a student must be skilled in algebra, differential equations, vector math, physics, and statics. The incorporation of multiple fundamental subjects offers a partial explanation of why exam grades in dynamics are lower than those for statics and thermodynamics (Froyd & Ohland, 2005). Difficult concepts of dynamics include force being related to acceleration—not velocity (Clement, 1982; White, 1983), vectors detailing both magnitude and direction, vector-math operating differently than scalar-math, frictionless conditions contradicting real-life experiences (White, 1983), and free-body diagrams representing complex objects and loadings (R. Streveler, Litzinger, Miller, & Steif, 2008). Furthermore, the topics of circular motion and pulleys challenge students (Fang, 2012; Viiri, 2003). Most, if not all, of these concepts pertain to knowledge assumed to be garnered from the prerequisite courses for Dynamics. Accordingly, Gray et al. (2005) concluded that "student misconceptions [of dynamics] are not random, but are generally the result of a deficiency in their understanding of fundamental principles" (p. 1). Therefore, the development of a student's conceptual understanding of fundamental dynamics principles is an important aspect of being successful in Dynamics.

Overall Research Approach

Research Methods

The research designs employed in this dissertation investigate the experiences and performance of specific subgroups of students. A pragmatic worldview undergirds the research. Multiple types of data and research methods are employed to better understand the students' performance on the aDCI and their resource-usage behaviors. Chapter 2 uses many quantitative methods (correlation, confirmatory factor analysis, item response theory, multiple-group confirmatory factor analysis) but relies on qualitative content analysis to understand why certain items on the aDCI may be gender biased. Chapter 3 utilizes an embedded research design that leverages quantitative (cluster analysis) and qualitative (thematic analysis) methods to understand how and why the students exhibited certain resource-usage behaviors. Chapter 4 uses quantitative methods (multiple regression analysis) to relate the students' resource-usage patterns to their achievement in Dynamics but contextualizes the statistical findings with the qualitative results from Chapter 3. In all three chapters, the results of the quantitative and qualitative analyses are treated equally when making inferences from the data because it is assumed that the qualitative and quantitative data complement one another. The strengths of qualitative data offset the weaknesses of quantitative data, and vise-versa. Overall, the common theme across all research designs is the use of analysis techniques that disaggregate the data to investigate the experiences and academic achievements of subgroups of students that would have gone unnoticed had I only analyzed the data in aggregate.

Research Setting and Context

Dynamics, which is formally known as ME 274: Basic Mechanics II, is a sophomore-level undergraduate engineering course at Purdue University taught by the School of Mechanical Engineering. Purdue University is a large, Midwestern, research-focused institution, and Mechanical Engineering is the largest school (department) within the College of Engineering. Dynamics is a three-credit-hour class that normally meets three times a week for 50 minutes at each meeting. It is typically taught during all three academic semesters at Purdue University (fall, spring, and summer), with the highest enrollment occurring in the spring semester. During the spring semesters, roughly 400 students are distributed across four sections. In the fall and summer, the number of students enrolled is much smaller—approximately 120 students across two sections in the fall and 40 students in one section for the summer. Instructors for the course vary in instructional experience from being first-time instructors to being a developer of Freeform.

The structure of Dynamics is largely uniform across sections. Instructors choose the pedagogical methods of their choice for classroom instruction, but all of the instructors to date have devoted the majority of the lecture time to solving example problems. The homework assignments and exams are common across all of the sections. The students submit two homework problems three times a week (for a typical Monday, Wednesday, Friday class), and, frequently, only one of those two problems is graded. The homework assignments consist of custom-written questions that change from semester-to-semester to discourage academic dishonesty. Solution videos for each homework problem are posted to the course website shortly after the homework due date.

The course utilizes three midterm exams and a comprehensive final exam. About two-thirds of each exam focuses on problem-solving skills and the remaining portion focuses on conceptual understanding. Together, the exams constitute 75% of a student’s grade. Homework performance usually counts as 17% of a student's grade, and scores on quizzes and a participation-based fundamentals exam make up the remaining 8% of a student’s grade.

Participants

A range of students from different majors and levels (e.g., sophomore, junior, senior, etc.) enroll in Dynamics, but the majority of students are sophomore-level, mechanical engineering students for whom Dynamics is a required class. For illustrative purposes, the demographic information from Purdue's Registrar office for the students who enrolled in Dynamics from Spring 2015-Spring 2018 is shown in Table 1. The demographic categories in Table 1 reflect how the Registrar collected the data. I acknowledge that the binary gender variable (that uses sex labels) is a simplification of the gender spectrum, and I recognize that the "ethnicity" variable confounds race, ethnicity, and international status. Approximately 20% of students who enrolled in Dynamics were international students, and all of those students chose "international" as their ethnicity. Therefore, the other racial/ethnicity groups listed in Table 1 describe the race/ethnicity of domestic students. The demographic characteristics of the sample used in each individual study of this dissertation are presented in their respective chapters.

The sample for a given study depends on data availability and if a student consented to participate in this research. The Purdue Institutional Review Board has approved post-hoc consent for all of the students who took the course prior to Fall 2015, and individualized consents have been gathered in all of the semesters since (and including) Fall 2015. The consent forms used for this study are included in Appendices A and B, and the surveys used to collect resource-usage and motivational data from the participants are included in Appendices C and D.

Table 1. Demographics of students who enrolled in Dynamics from Spring 2015-Spring 2018 (*N* = 2037).

Variable	%
Major	
Mechanical Engineering	78%
Agricultural Engineering	5%
Nuclear Engineering	5%
Multidisciplinary Engineering	4%
Other	7%
Ethnicity (Race/Ethnicity/ International Status)	
Domestic, White	58%
Domestic, Asian	7%
Domestic, URM	5%
Domestic, Other	6%
International	24%
Gender	
Male	82%
Female	18%

Note. The sum of the percentages for the major category does not equal 100% because of numerical rounding.

Researcher Positionality

I am a white man pursuing a doctorate in engineering education who has neither taught Dynamics nor experienced (Purdue's specific version of) Dynamics as a student. I have a bachelor's and master's degree in mechanical engineering, professional experience in engineering, and teaching experience at the university level. Therefore, while an outsider to Dynamics (although I have been researching it for three years) and women's experiences in engineering, I am an insider to the culture of engineering education. For my research involving gender, I collaborated with many experts in gender studies and validation research to ensure the quality of my work.

Chapter Organization

Three interrelated studies comprise this dissertation. All three studies are linked by their use of analysis methods that disaggregate the data to better understand the experiences of unique subgroups of students. Chapter 2 presents a validation study of the aDCI and subdivides the data by gender to investigate differential item functioning (DIF) and gender bias. Chapter 3 describes a study that uses quantitative methods to cluster, or group, students by their holistic (overall), resource-usage behaviors, and then uses qualitative methods to better understand how and why the students in each cluster used the resources as they did. Chapter 4 uses multiple regression analysis to investigate the academic achievement of students in each of the unique resource-usage clusters and draws from the qualitative work from Chapter 3 to make meaning of the regression results. One of the achievement measures considered in Chapter 4 is the students' performance on conceptual exam questions, including the aDCI questions evaluated in Chapter 2. Brief summaries of the three studies that constitute Chapters 2-4 are provided in the following sections.

Chapter 2. A Validation and DIF Study of the aDCI

Chapter 2 presents an investigation into the quality of the aDCI. A high-quality assessment must be valid, reliable, and fair (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). I used a variety of statistical methods, including confirmatory factor analysis (CFA) and item-response theory (IRT), to evaluate the validity and reliability of the aDCI. To my knowledge, this is the extent to which all of the previous validation studies of engineering CIs have been completed, representing lost opportunities to disaggregate the data and investigate the fairness of the other CIs across subgroups of students. I used multiple-group confirmatory factor analysis to evaluate the gender bias of the aDCI. The results of these analyses indicated that one problem was that the aDCI was measuring an unintended construct and that two other items were slightly biased against women. The gender bias of the two items likely adversely affected women's performance on the Dynamics final exam and, subsequently, their course grade, which highlights the practical consequences that unfair biases can have on certain subgroups of students. In Chapter 4, I excluded the scores of the three items with questionable validity or fairness when calculating a student's overall performance on conceptual exam questions to examine the impact of this bias on students' achievement in Dynamics.

Chapter 3. Identifying and Understanding Resource-Usage Patterns

The studies that are detailed in Chapter 3 and Chapter 4 worked in concert to investigate the relationship between the Dynamics students' resource usage behaviors and their achievement in the course. The cluster analysis of Chapter 3 identified nine archetypical resource-usage patterns based on self-reported survey data on how frequently each student used nine help resources. Then, to better understand how and why students enacted their respective resource-usage patterns, I completed a thematic analysis of student interviews for each cluster. The resulting themes suggested that the reasons students used or did not use a given resource closely aligned with the expectations and values they had for the resource, as described in Makara and Karabenick's (2013) expectancy-value model for help source selection. While this study identified that students across all nine resource-usage patterns sought help in order to build understanding of the material, not just in an attempt to quickly find the correct answer, the actual academic achievements of the students in different clusters were not compared until the completion of the study described in Chapter 4.

Chapter 4. Relating Resource-Usage Patterns to Academic Achievement

Extending the work of Chapter 3, Chapter 4 investigated the relationship between a student's resource-usage pattern and their achievement in Dynamics. Achievement was operationalized as a student's overall course grade, their overall performance on the problem-solving exam questions, and their overall performance on the conceptual exam questions. A student's overall grade was an aggregated, weighted-sum of a student's performance on all of the Dynamics assignments, whether they were group or individual assignments. The exam scores were purely individualistic measures of a student's achievement. The problem-solving exam questions measured a combination of procedural and conceptual knowledge (Rittle-Johnson & Schneider, 2015), and the conceptual questions were designed to only assess a student's conceptual knowledge of dynamics. When calculating a student's performance on the conceptual questions across all of the Dynamics exams, the three aDCI questions identified as having validity or gender-bias concerns were excluded.

Multiple regression analysis was used to investigate the relationship between a student's resource usage and their achievement while controlling for many demographic, cognitive, and non-cognitive factors that can influence resource usage or achievement. The results suggest that on overage a student's resource-usage pattern was not predictive of achievement, thereby indicating that, in general, the students used different combinations of resources in diverse ways to earn similar achievements. However, the regression results also indicated that students who primarily relied on their lecturebook and peers for support performed better than their similar peers in other clusters, and those who rarely used their lecturebook had lower course grades. The qualitive themes from Chapter 3 regarding the resource-usage behaviors of the students in these two clusters were used to inform general suggestions on how all future Dynamics students might use the resources to improve their achievement in Dynamics.

Contribution

Beyond being three exemplary studies for why it is important to utilize analysis methods that disaggregate the data to investigate the experiences and academic achievements of specific subgroups of students, this dissertation is impactful to the engineering education research community in three respects. First, to my knowledge, the validation study of the aDCI is the first validation study of an engineering CI to consider bias and fairness. Most CI validations only consider construct validity and reliability with aggregated data, which can leave unfair biases unnoticed, as Chapter 2 illustrates. Second, I believe Chapter 2 is the first CI validation study to formally implement an argument-based approach. Thus, Chapter 2 provides a rigorous, yet flexible, validation framework for other researchers to adapt to other CIs. Third, Chapters 3 and 4 provide a template for a new approach to studying students' help-seeking behaviors. Chapter 3 is the first study of students' help-seeking behaviors to utilize model-based clustering on data (self-reported or observed) of face-to-face and online resources. Other help-seeking studies that have used cluster analysis do so according to the students' general help-seeking tendencies. Thus, Chapter 3 provides a flexible framework for instructors and researchers to identify archetypical resource-usage patterns of the students in a specific course that can combat stereotypes and misconceptions about the help-seeking behaviors of the average, or typical, student. Chapter 4 is the first study to consider the students' resource usage holistically when correlating it with academic achievement. Without disaggregating the data and considering holistic resource-usage patterns, the unique achievements of two, smaller subsets of students likely would have gone unnoticed. Overall, the knowledge of how students experience assessments and the resources of the course enables instructors to better coach students on how to be successful in the course and to modify the curriculum and resources of the course to better support the learning of all students.

CHAPTER 2. A VALIDATION AND DIF STUDY OF THE ADCI

Stites, N. A., Douglas, K. A., Evenhouse, D., Berger, E., DeBoer, J., & Rhoads, J. F. (in press). A validation and differential item functioning (DIF) study of an Abbreviated Dynamics Concept Inventory. *International Journal of Engineering Education.*

Abstract

Concept inventories (CIs) have become popular assessment tools in science, technology, engineering, and mathematics education. Some researchers use CI scores when looking at differences in conceptual understanding or learning gains across demographic groups, but very few CIs have been evaluated for measurement bias or other aspects that threaten the fair assessment of learners. The most common psychometric evaluation models are shaped primarily by the majority demographic group, so these models can hide biases in the assessment against minority groups. The purpose of this study was to evaluate the extent to which the validity, reliability, and fairness evidence supports the use of the total score on a 12-item Abbreviated Dynamics Concept Inventory (aDCI) as a measure of a student's overall conceptual understanding of dynamics. Because of the strong relationship between the aDCI and the Force Concept Inventory, which has previously been shown to include item-level gender biases, we examined threats to fair measurement across gender scores of the aDCI. We employed an argument-based validation approach which tested: 1) the fit of a single-factor latent structure for the aDCI scores via a confirmatory factor analysis (CFA), 2) the difficulty and discrimination of each item using item response theory, 3) the correlation between the aDCI scores and similar measures of conceptual understanding, and 4) the differential item functioning of the aDCI items across gender groups via a multiple-group CFA. We found that one item had face-level construct validity concerns and two others were slightly biased against women. Possible sources of gender bias included the question's content and context. Our results suggest that the interpretation of a student's total aDCI score should consider the differential item functioning of two items across gender and the construct-alignment concerns of a third item. This work highlights the importance and challenge of designing inclusive assessments and validating them with fair psychometric models.

Keywords: concept inventory, validity, reliability, fairness, gender, engineering education, assessment bias

Introduction

Research suggests that a student's conceptual understanding of fundamental engineering topics directly relates to their ability to solve problems and apply existing knowledge to new and novel situations (Hatano & Inagaki, 1984; McKenna, 2007; Pandy, Petrosino, Austin, & Barr, 2004; R. Streveler et al., 2008)[1]. Concept inventories (CIs) are increasingly-popular instruments for assessing students' conceptual understanding, as well as their misconceptions, within a particular domain (such as statics, dynamics, or thermodynamics; Jorion et al., 2015). The interest in CIs in engineering increased significantly in the early 2000s, potentially driven by a transition of ABET accreditation guidelines to a focus on program outcomes (Reed-Rhoads & Imbrie, 2008). Currently, the development and assessment of conceptual understanding is still a large endeavor; a search of the U.S. National Science Foundation awards found over $7 million in active awards with the phrase "concept inventory" in the proposal abstract alone. CIs are commonly used to evaluate pedagogical innovations (Freeman et al., 2014; Hake, 1998; Hestenes et al., 1992), and they have also been used to better understand how students develop conceptual understanding (Henderson, 2002). Yet, despite the investment in and positive outcomes associated with CI use, research on the quality and fair use of these assessment instruments is generally incomplete (Jorion et al., 2015).

Researchers have used many different types of evidence to validate the use of CIs, with varying degrees of quality (R. A. Streveler et al., 2011). Because validity pertains to justifying specific interpretations and uses of assessment scores, evidence must be collected to test the plausibility of the desired claims made from the scores (Messick, 1989). Generally speaking, developers and users of CIs have similar desired inferences from the CI scores–the students' conceptual understanding of a specific topic. Therefore, in response to the need for more consistency, researchers have begun to develop guidelines to aid those interested in developing

[1]Copyright permission from IJEE included in Appendix E

or using CIs for their research. Streveler et al. (2011) demonstrated how the Assessment Triangle can be applied to the development and testing of CIs, where evidence to support the interpretation of CI scores was empirically gathered through studies of item difficulty and discrimination. The Assessment Triangle provides a framework for assessment development that ensures the alignment between cognitive theory, observing the students' assessment responses, and interpreting the responses (National Research Council, 2001). Focusing on the interpretation corner, Jorion et al. (2015) suggested a framework to evaluate the plausibility of three common claims made from CI results: 1) students' overall conceptual understanding, 2) students' understanding of specific concepts, and 3) students' propensity for misconceptions. While these frameworks are helpful for developing and evaluating CIs, the examples do not consider use among diverse learners. According to the *Standards for Psychological and Educational Assessment*, high-quality assessments are based on evidence of reliability, validity, and fairness (American Educational Research Association et al., 2014). Fair assessment has received relatively little attention in engineering education with few examples of what is meant by "fair" and how to measure it. This work provides one example of how to operationalize and measure fairness and, to our knowledge, represents the first psychometric analysis of an engineering CI to consider fairness.

Psychometric models used in the validation of assessment instruments, such as CIs, are based on statistics for which the responses of the demographic majority group will have the most power in shaping the model. Given that only approximately 20% of U.S. engineering students are women (National Science Board, 2018), any psychometric model from that sample is essentially normed on the responses from men. To examine how the items perform for minority students, researchers need to purposefully examine measurement models for minority groups. In a recent review of assessment development articles published in engineering education journals, only one article considered potential bias in the assessment items themselves (Kerrie A. Douglas, Rynearson, & Purzer, 2016). Yet, recent research in engineering education assessment validation found items that had acceptable fit for the whole student group also contained item-level bias against English-language learners (Kerrie A. Douglas, Fernandez, Purzer, Fosmire, & Van Epps, 2018). An acceptable psychometric model fit for the whole group does not guarantee that those same items are fair for all students. The evaluation of test items is a prerequisite to the evaluation of the learners who responded to those items. Score differences found between groups cannot be justifiably interpreted as true score differences unless bias in the measurement model has been ruled out. It is simply unknown whether assessment questions or answer choices are understood in the same way by diverse groups of students, unless the evidence is specifically sought. Therefore, the evaluation of CIs in engineering for fairness across all underrepresented groups, including those based on gender, is a prerequisite for the fair use of the students' scores to make decisions of personal consequence to the students.

Purpose of the Study and Research Questions

The overall purpose of this research is to report the development and initial validation studies of a shortened form of the Dynamics Concept Inventory (DCI; Gray et al., 2005), which is named the Abbreviated Dynamics Concept Inventory (aDCI; Appendix F includes a full list of the abbreviations used in this paper). While the DCI is an established instrument used in engineering education research (e.g., Coller, 2015; Self & Widmann, 2017), a shortened version would enable instructors and researchers to assess students' conceptual understanding of dynamics in less time (Stites et al., 2016). The research question that guides this work is: to what extent does the validity, reliability, and fairness evidence support the use of aDCI scores as a measure of students' overall conceptual understanding of dynamics?

Regarding the fairness of the aDCI, we focus on gender fairness in this paper because the aDCI is closely related to the Force Concept Inventory (FCI) that is used in physics education, and the FCI has been shown to include item-level gender bias (Traxler et al., 2018). Additionally, previous research has indicated a statistically-significant gender gap in the students' total scores on the aDCI (Prebel et al., 2017).

In accordance with Messick's (1990) description of validation research as hypothesis testing and Kane's (1992) argument-based approach to validation, we investigate the overarching research question by testing the following hypotheses:

If a student's total score on the aDCI can be interpreted as a measure of their overall conceptual understanding of dynamics, then:

Hypothesis 1. a single-factor latent structure would effectively model the shared variance of the aDCI items;

Hypothesis 2. the aDCI items would be appropriately difficult and able to discriminate between students with high and low overall conceptual understanding of dynamics;

Hypothesis 3. the aDCI total score would be correlated to similar measures of overall conceptual understanding of dynamics;

Hypothesis 4. The aDCI items would function similarly for students of equal ability regardless of their background or socialization, including gender.

The purpose of each hypothesis and the analytical methods used to investigate the hypotheses are summarized in Table 2. The structure of this study follows the order of the hypotheses. The results of Hypotheses 1-3 provide information regarding the reliability and construct validity of the aDCI scores, and Hypothesis 4 targets fairness.

Table 2. An overview of the analytical methods used in this study and the purpose of each of the hypotheses.

Hypothesis	**Analytical Method**	**Purpose**
1. Single Factor Latent Structure	Confirmatory Factor Analysis (CFA)	Determine if all items of the aDCI serve as indicators of a single latent construct that is assumed to be a student's overall conceptual understanding of dynamics
2. Appropriate Difficulty and Discrimination	Item Response Theory (IRT)	Investigate how well the difficulty of the aDCI items match the latent abilities of the students and how well the items differentiate the higher- and lower-performing students
3. Correlated to Similar Measures	Correlation	Evaluate the relative relationships between similar measures of students' overall conceptual understanding of dynamics
4. Measurement Invariance Across Groups	Multiple-Group Confirmatory Factor Analysis (MG-CFA)	Determine if the aDCI functions the same for men and women; i.e., evaluate the aDCI for gender bias

Literature review

Reliability, Validity, and Fairness

The cornerstones of high-quality assessments reside in the evidence of reliability, validity, and fairness (American Educational Research Association et al., 2014). Reliability refers to the degree of consistency both internal to the assessment and of the scores for multiple administrations of the assessment (Kerrie A. Douglas et al., 2016). Validation is the process of identifying multiple sources of relevant evidence to make a judgement about the appropriateness of using a given assessment for a specific purpose (Kerrie A. Douglas & Pellegrino, 2017; Kerrie Anna Douglas & Purzer, 2015). Thus, validity refers to the evidence and rationale for claiming an assessment score can be interpreted and used as intended – as a measure of the learners' knowledge, skill, or conceptual understanding (Kerrie A. Douglas & Pellegrino, 2017; Kerrie A. Douglas et al., 2016; Messick, 1989). Of the three cornerstones of high-quality assessments, validity is overarching. Validity depends on the evidence of reliability and fairness; for an assessment to have a valid use, it must first demonstrate reliability and fairness in assessing learners.

There is no one set of procedures for validation because the validation process depends on the specific interpretation and purpose of the assessment (Mislevy & Haertel, 2006). In order to holistically evaluate the use of an assessment, one would clearly articulate the chain of reasoning involved in determining what evidence to test (Mislevy & Haertel, 2006). In the case of concept inventories used in physics or engineering education, after the assessment is administered, validity testing would begin with "If this assessment score truly measures the students' conceptual understanding, then what else has to be true so that the reliability, validity, and fairness evidence supports this argument?"

While most engineering education researchers are at least aware of the terms "reliability" and "validity" in educational assessment, fairness is less understood. Fairness was recently raised to the same level of importance as validity in the *Standards for Educational and Psychological Testing* in order to emphasize how crucial evidence of fairness is for ethical education assessment (American Educational Research Association et al., 2014). The term itself, fairness, does not have one specific technical meaning, as it has been used in a variety of ways in educational assessment (American Educational Research Association et al., 2014). The

Standards identify common views of fairness to include equitable treatment during the testing process, lack of measurement bias, access to content assessed, and valid interpretations of individual test scores. Fair and valid interpretations of test scores can depend on, among other factors, the content assessed and the context of the questions (American Educational Research Association et al., 2014; Ding & Caballero, 2014). Measurement bias and valid interpretation of individual test scores are the most pertinent views of fairness for this work because they are partially dependent on item-level bias that can cause differential item functioning (DIF) across student groups, which is what we investigate in Hypothesis 4. Researchers have previously found item-level gender bias in physics CIs (e.g., Dietz et al., 2012; Osborn Popp, Meltzer, & Megowan-Romanowicz, 2001; Traxler et al., 2018) and physics (mechanics) is closely related to dynamics. Therefore, we investigate threats to the gender fairness of the aDCI stemming from the psychometric models of evaluation and from the content and context of the questions.

Sources of Gender Bias in CIs

Because of the minimal research on the fairness of engineering CIs, we looked to the literature from physics education research for information regarding possible sources of gender bias in CIs. Madsen, McKagan, and Sayre (2013) reviewed literature on the gender gap of physics concept inventories, and they identified six categories of factors that had evidence of a demonstrated impact on the gender gap: background and preparation (e.g., high school background), gender gaps on other measures (e.g., average exam scores), differences in personal beliefs and the answer a "scientist" would give, teaching method (e.g., level of interactive engagement), stereotype threat, and question wording.

Regarding question wording, the conclusion of Madsen and colleagues was largely based on McCullough's findings (2004, 2011) that students changed how they answered questions on the FCI when the question wording was revised to included everyday and stereotypically feminine contexts (rather than stereotypically masculine contexts of the traditional FCI). However, the way in which the context influenced the students' performance on individual questions was inconsistent, meaning the gender gap for the overall scores remained unchanged for McCullough's revised concept inventory. Nonetheless, McCullough's findings showed that changing the context of an individual question affects how men and women answer the question.

McCullogh's findings aligned with what Ding and Caballero (2014) called a *context* effect. A context effect is when one group of students is more familiar with the non-essential features of a question (such as wording, language, or images), and this extra familiarity with the context causes DIF. Alternatively, Ding and Caballero posited that DIF could be caused by a *content* effect, which is when groups of students who have been exposed to different interventions, instruction, or experiences perform differently on an item. Unfamiliar content and contexts can create extra cognitive load which can affect a student's performance because the student must first infer the situation described in the problem statement before they can attempt to solve the problem (Rennie & Parker, 1993). Thus, content and context effects can favor certain groups based on their background and socialization, including gender.

To help instructors identify and eliminate gender bias in physics questions, Rennie and Parker (1993) developed a framework, see Table 3, for assessing the gender orientation (masculine, feminine, allegedly neutral, or gender inclusive) of physics questions along four dimensions (language, portrayal of stereotypes, appeal to background experiences, and context). Later, McCullough (2004, 2011) used the same framework to categorize the items of the FCI. Leveraging the strong relationship between physics and dynamics, we used this framework to qualitatively evaluate the aDCI items for gender bias.

Table 3. A gender-orientation framework for evaluating items on the aDCI for gender bias (Rennie & Parker, 1993).

Criteria	Masculine Orientation	Feminine Orientation	Allegedly Neutral Orientation	Gender-Inclusive Orientation
Language	Uses he, him, his	Uses she, her, hers	Uses they, them, their Uses role (e.g., a sprinter…)	Uses the name of a person Uses "you"
Portrayal of Stereotypes	Men in active roles, women in passive roles	Women in active roles, men in passive roles	Genderless people in active roles (e.g., a scientist…)	Both men and women in active and passive roles
Appeal to background experiences	Relevant to stereotyped experiences of men	Relevant to stereotyped experiences of women	Not relevant to human experiences	Relevant to men and women equally
Context	Decontextualized, abstract	Human, social	Concrete setting	Human, social, environmental

Note. Rennie and Parker used the terms *male* and *female* rather than the terms *masculine*, *feminine*, *men*, and *women* (as shown). Rennie and Parker included the word "allegedly" to the Neutral Orientation category because their research indicated that students assume plural pronouns and genderless people refer to men.

Background

The sophomore-level dynamics course required by many engineering majors is often challenging. It is a gateway course to the more specialized upper-division engineering courses, and, when paired with statics, it creates the problem-solving and conceptual foundation for much of the curriculum in many engineering disciplines. To be successful in dynamics, a student must understand algebra, differential equations, vector math, physics, and statics. The incorporation of so many fundamental subject areas of engineering may be a partial explanation of why students' exam scores for dynamics courses are lower than they are in statics and thermodynamics courses (Froyd & Ohland, 2005). Many researchers have discussed the difficult aspects of dynamics (e.g., Clement, 1982; Shryock & Froyd, 2011; R. Streveler et al., 2008; White, 1983), many of which involve prerequisite material. The difficulties that the students have with the prerequisite fundamentals support the conclusion of Gray et al. (2005) that "student misconceptions are not random, but are generally the result of a deficiency in their understanding of fundamental principles" (p. 1). Accordingly, Cornwell (2000) noted that when students do not understand the fundamentals of dynamics, they struggle to identify when or why to apply a given model or solution approach.

To help instructors assess their students' conceptual understanding of the fundamental topics of dynamics, Gray and colleagues (2005) developed the Dynamics Concept Inventory (DCI). The DCI stemmed from the need to quantitatively assess the efficacy of pedagogical innovations in dynamics. Gray et al. conducted a modified Delphi process, focus groups, student interviews, (informal) instructor interviews, and pilot tests to develop the DCI. The final result was a 29-item instrument that targeted 11 of the most important and difficult concepts in dynamics (Gray et al., 2005). Each item included five answer choices. Psychometric analyses have found that the DCI should be used for low-stakes assessment and that the total scores could be interpreted as the students' overall understanding of concepts on the DCI (Gray et al., 2005; Jorion et al., 2015). Thus, it is plausible that a carefully-selected subset of DCI items could provide a similar measure of the students' conceptual understanding.

aDCI Development

To streamline the implementation of a dynamics CI and to save class time (Stites et al., 2016), a shortened version of the DCI (the aDCI) was developed and incorporated into the final exam of a dynamics course. We note that Jorion et al.'s (2015) suggestion of using the DCI in a low-stakes environment was not yet published. The number of conceptual questions on past final exams for this dynamics course typically ranged from 5 to 13. Therefore, the goal for the aDCI was to target as many of the important and difficult dynamics concepts as possible with fewer than 13 items.

The DCI developers did not specify which items targeted which concepts, and very limited psychometric information was available for the DCI at the time that the aDCI was developed (early 2015). Therefore, two of the co-authors of this paper (both subject-matter experts in dynamics) used their best judgement to categorize the DCI items according to conceptual content. They then chose 11 items for inclusion in the aDCI that spanned 10 of the 11 conceptual categories and a twelfth item that tested pre-requisite physics knowledge. The questions were selected based on clarity and alignment with the material taught in the dynamics course, which reflected the curriculum of most undergraduate dynamics courses and included the study of particle and rigid-body kinematics and kinetics in two and three dimensions. The twelve selected items for the aDCI and their targeted concepts are listed in Table 4.

Table 4. Description of the concepts assessed by each item of the aDCI (using verbatim descriptions from Gray et al. (2005)).

aDCI Item #	Concept Description
Q1	Newton's third law dictates that the interaction forces between two objects must be equal and opposite.
Q2	Angular velocities and angular accelerations are properties of the body as a whole and can vary with time.
Q3	If the net external force on a body is not zero, then the mass center must have an acceleration and it must be in the same direction as the force.
Q4	In general, the total mechanical energy is not conserved during an impact.
Q5	An object can have (a) nonzero acceleration and zero velocity or (b) nonzero velocity and no acceleration.
Q6	The direction of the friction force on a rolling rigid body is not related in a fixed way to the direction of rolling.
Q7	The angular momentum of a rigid body involves translational and rotational components and requires using some point as a reference.
Q8	If the net external force on a body is not zero, then the mass center must have an acceleration and it must be in the same direction as the force.
Q9	The inertia of a body affects its acceleration.
Q10	A particle has acceleration when it is moving with a relative velocity on a rotating object.
Q11	Points on an object that is rolling without slip have velocities and acceleration that depend on the rolling without slip condition.
Q12	Different points on a rigid body have different velocities and accelerations, which vary continuously.

Note. Q3 and Q8 assess the same concept.

Methods

Participants and Data Collection

The aDCI data for this study were collected from students enrolled in a sophomore-level dynamics course at a large, public, doctoral university with the highest category of research activity (Indiana University Center for Postsecondary Research, n.d.) located in the Midwest region of the United States. The dynamics course was focused on particle and rigid-body kinematics and kinetics, as well as mechanical vibration. Each year, over 500 students enrolled in the course, often in class sections of up to 120 students. The sampling frame for this study consisted of all of the students who enrolled in the course from Spring 2015-Spring 2017. Of the 1,397 students in the sampling frame, 1,351 students completed the aDCI, and 1,250 of those students agreed to participate in the research study. The aDCI was administered as part of the course's final exam, and the items were scored as correct or incorrect (1 or 0, respectively). If an item was unanswered or if multiple answers were selected (which occurred 0.31% of the time), the response was considered incorrect. These scoring methods led to a sample with no missing data.

The demographic characteristics of our sample are shown in Table 1. The institutional-research data we used conflated race, ethnicity, and international status into one variable and collected gender as a binary variable (which we acknowledge is a simplification of the gender spectrum). The proportion of women in this course is representative of many mechanical engineering courses at large research universities in the USA, including those at the university of this study.

Table 5. Demographics of the sample (N = 1250).

Variable	Value	
Major[a]	81%	Mechanical Engineering
	4%	Nuclear Engineering
	5%	Agricultural Engineering
	3%	Multidisciplinary Engineering
	6%	Other
Race/Ethnicity/ International Status	60%	Domestic, White
	7%	Domestic, Asian
	5%	Domestic, URM
	23%	International
	5%	Domestic, Other
Gender	82%	Male
	18%	Female

[a]The total percentages of major does not sum to 100% because of numeric rounding.

Data Analyses

Preprocessing Data: Descriptive and Correlation Statistics

Prior to testing the four psychometric hypotheses, the data were explored via descriptive statistics and correlations. Because of the dichotomous nature of the data (0 = incorrect, 1 = correct), the proportions of students who answered an item correctly and inter-item tetrachoric correlation coefficients were calculated (Scott, Schumayer, & Gray, 2012). The proportions provided a measure of item difficulty (Tavakol & Dennick, 2011); the tetrachoric correlations were measures of internal reliability and how related the items were to one another (Scott et al., 2012).

Hypothesis 1: A Single-Factor Latent Structure

This analysis used confirmatory factor analysis (CFA) to evaluate the hypothesis that the aDCI scores reflect a unidimensional latent-factor structure, i.e., conceptual understanding of dynamics. To identify the model and estimate all the factor loadings, the variance of the latent variable was constrained to be unity. A weighted least squares estimator in the *lavaan* package (version 0.5-23.109) of R (version 3.3.2) used diagonally weighted least squares to estimate the model parameters, and it used the full weight matrix to compute robust standard errors and a mean- and variance-adjusted chi-squared (χ^2) statistic. The estimator specified the model parameters that most accurately reproduced the tetrachoric correlation matrix for the sample data.

We holistically evaluated the model through the goodness of fit statistics of χ^2, comparative fit index (CFI), and root-mean-square error of approximation (RMSEA) goodness of fit statistics. We gave the statistical significance of the χ^2 test statistic minimal consideration when determining overall model fit because of its sensitivity to sample size and non-normality (Cheung & Rensvold, 2002; Hu & Bentler, 1998). More weight was given to the CFI and RMSEA values. As suggested by Hu and Bentler (1999), we considered CFI values above 0.950 and RMSEA values below 0.050 to be indicators of good model fit.

Hypothesis 2: Items of Appropriate Difficulty and Discrimination

Item response theory (IRT) models the probability of a student answering an item correctly as a function of their ability level (a latent trait) and the properties of the item that are independent of the sample. Similar to CFA, IRT utilizes a single-factor model to estimate each student's latent ability, which we again assumed to represent a student's overall understanding of dynamics. We used a 3-parameter (3PL; difficulty, discrimination, and guessing) model to characterize each item of the aDCI. The proportion of lower-performing students (those with an aDCI total score of 3 or less) who answered the item correctly was used as the initial value for the guessing parameter in the IRT model. The M2 test statistic was used to evaluate model fit, using $p < 0.050$ as the significance threshold (Maydeu-Olivares & García-Forero, 2010). The items' difficulty values were compared to the students' ability levels with a Wright map (Jorion et al., 2015) to determine if questions were too challenging or easy for our sample.

Discrimination values indicated how well the item differentiated students who knew the concept and those who did not (Tavakol & Dennick, 2011).

Hypothesis 3: Correlation with Similar Measures of Conceptual Understanding

Every intermediate exam in the dynamics course in which our participants were enrolled included conceptual questions, and in aggregate, the concepts assessed by the intermediate exams reflected the concepts assessed by the aDCI. Strong correlations between a student's performance on the conceptual questions of the three intermediate exams, their total score on the aDCI, their latent factor score from the CFA (from Hypothesis 1), and their ability score from the IRT analysis (from Hypothesis 2) would support the assumption that these data were all measures of the students' overall conceptual understanding of dynamics. The exact concepts assessed on the intermediate exams varied slightly across semesters. To be able to compare the instructor-written questions across semesters, we standardized the students' scores for each semester individually. Regarding format, the aDCI consisted of multiple-choice questions only, and the intermediate exams incorporated multiple-choice, true/false, and short-answer conceptual questions.

Hypothesis 4: Measurement Invariance Across Genders

If assessment items are truly measuring the intended construct and not outside factors, there should be no group level differences in item performance. Measurement invariance refers to the assumption that the measurement model is not significantly different for different demographic groups (Schmitt & Kuljanin, 2008). Conversely, differential item functioning (DIF) occurs when an item functions differently for different demographic groups (Bauer, 2017). There are multiple methods that can be used to detect DIF including multiple-group confirmatory factor analysis (Vandenberg & Lance, 2000), IRT techniques (Stark, Chernyshenko, & Drasgow, 2006), and non-parametric techniques like the Mantel-Haenszel method (Socha, DeMars, Zilberberg, & Phan, 2015). We used multiple-group confirmatory factor analysis (MG-CFA) for this study so that we could test the invariance of the relationships between the items and the latent variable (the factor loadings) and the item thresholds (the probability that a student will answer the item correctly) independently. The testing of measurement invariance with MG-CFA involves simultaneously fitting separate measurement models (with the same latent structure) to the data from men and women. Then, differences in the parameter estimates (such as factor loadings and thresholds) across the two measurement models are investigated by sequentially adding equality constraints to the parameter estimates of both models while testing for statistically significant changes in the fit of the overall model (which includes the measurement models of both men and women).

Brown (2015) referred to four levels of increasingly strict measurement invariance as: equal form, equal factor loadings, equal thresholds, and equal indicator residuals. We only tested equal form, equal factor loadings, and equal thresholds because the variances of the indicator residuals were calculated values, not estimated parameters, for our data type. For testing equal form, we compared the goodness of fit statistics and factor loadings for CFA models that used data from men only, women only, and men and women simultaneously but in separate factor structures (which we labeled Model 1). To test for equal factor loadings (Model 2), we constrained the unstandardized factor loadings for each item, respectively, to be equal across gender groups. Equal factor loadings indicate that the relationships between the items and the latent factor are the same for men and women (Cheung & Rensvold, 2002; Kline, 2016; Meredith, 1993).

The testing of equal thresholds for all of the items in aggregate (Model 3) incorporated the CFA assumption that a continuous, normally-distributed variable underlies the dichotomous score for each indicator. The threshold corresponds to the z-score that bisects the distribution curve such that the areas under the curve correspond to the proportions of students answering the questions as 0 or 1. Measurement invariance at the equal-thresholds level indicates that, on average, the items are not biased against either of the gender groups (Cheung & Rensvold, 2002; Kline, 2016; Meredith, 1993).

Because the fit statistics used to judge measurement invariance indicated how well the model reproduced the variances and covariances of the sample data overall (and in aggregate), the test of equal thresholds for all of the items simultaneously could hide biased thresholds for individual items. The test for equal thresholds for individual items was a two-phase process. First, we iteratively and individually released the equality constraint for each item's threshold to determine the statistical significance ($\Delta\chi^2$ *p*-value) of the change in the model fit when compared to Model 3. Second, we sequentially incorporated as many of the unequal thresholds as necessary into a final MG-CFA model. The second phase used a sequential model-improvement procedure similar to that used when altering a model based on modification indices (Tabachnick & Fidell, 2013, p. 733); the baseline model was updated whenever a model with a newly-released threshold constraint fit the data better than the existing baseline model. Our modification indices (values that are used to rank model changes according to how likely the changes are to improve the model fit) were the $\Delta\chi^2$ *p*-values from the first phase. The thresholds for the item with the lowest $\Delta\chi^2$ *p*-value from the initial phase were freely estimated first, and the resulting model fit was compared to that of the baseline MG-CFA model. Then, the same testing process was repeated for the item with the second-lowest $\Delta\chi^2$ *p*-value, then the third-lowest, and so on, identifying a new baseline each time the release of a threshold constraint resulted in a statistically significant model-fit improvement.

The same goodness of fit indices (χ^2, CFI, and RMSEA) and their thresholds used in the prior CFA were used for the measurement invariance tests. When nested models were compared, we used a χ^2 difference ($\Delta\chi^2$) test and the change in CFI to judge if the model fit changed significantly. Because the WLS estimator adjusts the test statistic for mean and variance, a scaled $\Delta\chi^2$ test according to Satorra's method (Satorra, 2000) was utilized. If the *p*-value for a scaled $\Delta\chi^2$ test was lower than 0.050, we rejected the null hypothesis of equivalent model fits. We considered a change in CFI greater than 0.010, as suggested by Cheung and Rensvold (2002), indicative of significantly different model fits.

Results

Descriptive and Correlation Statistics

The proportion of students answering each of the 12 items correctly, the inter-item tetrachoric correlation coefficients, and the item-test correlation coefficients (a measure of discrimination) are shown in Table 6. The low inter-item correlation coefficients illustrated the broad, and in some cases independent, nature of the concepts assessed on the aDCI, but Q7 had particularly low inter-item correlations, which was evident in our CFA as well. The lack of groups of items with high correlations suggested that a one-factor latent structure was the most probable model.

Table 6. Low correlation coefficients (lower diagonal with standard errors in the upper diagonal) between items of the aDCI illustrated the broad nature of the concepts assessed by the aDCI.

		Correlation Coefficients											
Item #	Proportion Correct	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1	0.91		0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07
Q2	0.80	**0.40**		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
Q3	0.83	0.22	**0.35**		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
Q4	0.63	0.18	0.28	0.29		0.04	0.04	0.05	0.04	0.05	0.05	0.04	0.05
Q5	0.63	0.21	**0.38**	0.27	0.27		0.04	0.05	0.04	0.05	0.04	0.04	0.05
Q6	0.45	0.23	0.27	**0.38**	0.21	0.24		0.05	0.04	0.05	0.04	0.04	0.05
Q7	0.67	0.07	0.12	0.18	0.06	0.13	0.18		0.05	0.05	0.05	0.05	0.06
Q8	0.41	0.18	0.19	0.07	0.25	0.26	0.14	0.08		0.04	0.04	0.04	0.05
Q9	0.36	0.26	0.23	0.07	0.21	0.21	0.12	0.17	0.28		0.04	0.04	0.05
Q10	0.42	0.13	0.25	0.08	0.13	0.26	0.21	0.19	0.18	0.20		0.04	0.05
Q11	0.66	0.22	**0.36**	0.20	0.28	0.22	0.24	0.15	0.26	**0.32**	**0.38**		0.05
Q12	0.84	0.12	0.24	0.22	0.22	0.22	0.18	0.09	0.10	0.15	0.13	0.23	
Total Score	0.63	0.32	0.49	0.40	0.47	0.50	0.47	0.36	0.44	0.45	0.46	0.52	0.34

Note. The correlations coefficients greater than 0.30 are bolded. A correlation coefficient of 0.30 indicates that approximately 10% of the variance in one item is explained by the variance in the other item (Cohen, Cohen, West, & Aiken, 2015, p. 38). N = 1250.

Table 7. Goodness of fit and model comparison statistics for testing measurement invariance of the aDCI across men and women.

		Overall Model Fit Indices						Change in Fit Indices			
Model #	Model Description	df	χ^2	χ^2 *p*-value	RMSEA (90% Conf. Interval)[a]	CFI	Comparison	Scaled $\Delta\chi^2$	Scaled df	Scaled $\Delta\chi^2$ *p*-value	ΔCFI
CFA for All Participants in Aggregate (12 items)											
-	Men and Women	54	90.08	0.002	0.028 (0.021, 0.036)	0.954	-				
Overall Measurement Invariance (11 items, Q7 removed)											
-	Men	44	70.37	0.007	0.030 (0.020, 0.039)	0.950	-				
-	Women	44	39.52	0.664	0.023 (0.000, 0.052)	0.974	-				
1	Equal Form	88	109.89	0.057	0.028 (0.018, 0.038)	0.954	-				
2	Equal Factor Loadings	99	124.91	0.040	0.023 (0.011, 0.033)	0.965	1 vs. 2	1.80	2.34	0.482	0.011
3	Equal Thresholds	109	143.74	0.014	0.025 (0.015, 0.034)	0.954	2 vs. 3	4.13	2.40	0.172	-0.011
Evaluation of Equal-Threshold Invariance for Selected Items											
4	Q3 Thresh. Est.	108	137.16	0.030	0.023 (0.000,0.052)	0.960	4 vs. 3	2.42	0.34	0.032	0.006
5	Q3, Q6 Thresh. Est.	107	131.58	0.054	0.028 (0.018,0.038)	0.965	5 vs. 4	1.66	0.27	0.044	0.005
6	Q3, Q6, Q4 Thresh. Est.	106	130.31	0.055	0.023 (0.011,0.033)	0.966	6 vs. 5	0.47	0.30	0.164	0.001

Note. n_{men} = 1031, n_{women} = 219. "Thresh. Est." indicates that an item's threshold was freely estimated across gender groups. χ^2 = chi-squared fit statistic with robust errors; df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index. Chi-squared difference tests for nested model utilized the Satorra 2000 for scaling the chi-squared statistic and df.

[a]The *p*-values for all RMSEA values listed in this table were greater than 0.990, except for the CFA for women only which was greater than 0.930.

Hypothesis 1: A Single-Factor Latent Structure

The goodness of fit statistics for the model with a single-factor latent structure are included in Table 7 (which includes results from Hypothesis 4 also). The CFI was above the threshold for considering the model a good fit (0.950), and the RMSEA was below its 0.050 threshold. The χ^2 value had an associated *p*-value of less than 0.050, but this was not surprising given the dependence of the χ^2 statistic on sample size (Cheung & Rensvold, 2002). Therefore, the evidence suggests that a single-factor latent model, as shown in Figure 1, fits the data well and supports the hypothesis that all items were indicators of a single latent construct—the overall conceptual understanding of dynamics.

Additionally, all of the factor loadings for the model shown in Figure 1 were statistically significant ($p < 0.001$). The factor loading for Q7, however, was nearly half that of the factor loadings for the other items. This low factor loading indicates that Q7 may be measuring a different construct than that measured by the other 11 items, and potential causes of the psychometric properties of Q7 are explored in the Discussion section.

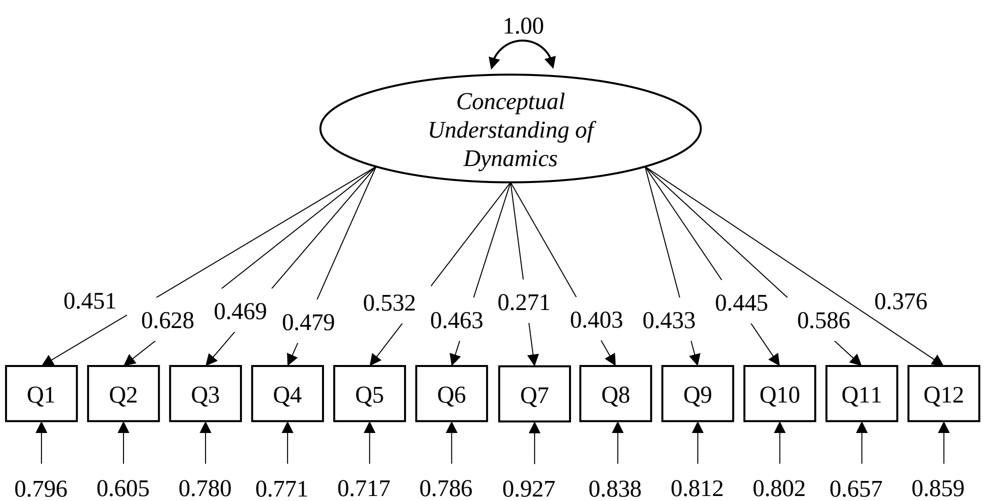

Figure 1. The single-factor structural model fit the aggregated data from men and women well, as tested in Hypothesis 1. The numbers on the arrows from the latent construct (conceptual understanding of dynamics) to the items (Q1-Q12) represent factor loadings, which in this study are equivalent to correlation coefficients. The numbers below the items indicate the proportion of unexplained variance in each item.

Hypothesis 2: Items of Appropriate Difficulty and Discrimination

The M2 test statistic for the 3PL model ($M2$ = 55.79, df = 42, p = 0.075) indicated that there was no statistically significant difference between the observed data and the model-fitted data. Three conclusions can be drawn from the item characteristic curves in Figure 2 and the parameter values in Table 8.

First, all items (except for Q7) had a positive and relatively high discrimination (maximum slope steepness) between the ability levels of -2 and 2, which was the ability range of the students in our sample. The lower discrimination value of Q7 (which is represented by the shallower slope in Figure 2) means that Q7 did not efficiently differentiate the high- and low-ability students based on their response.

Second, Q1 and Q3 had difficulty values near negative two, which were considerably lower than most of the other questions. As shown in Figure 3, Q1 and Q3 were most suited to differentiate students at a low ability level (near -2), and our sample had very few students with such low ability. The y-axis of Figure 3 shows the logit transformations of the item difficulties and the students' abilities on the same scale (Wilson, 2005). It is preferable to have the question difficulties in the ability range with the highest density. While the power to differentiate students in our sample would have improved with higher difficulty levels for Q1 and Q3, it was expected that the students would perform well on these items because these items assessed less-challenging, prerequisite content.

Third, most of the items had non-zero guessing parameters, and many were above what would be expected for random guessing (0.20) on items with five possible answers. Therefore, the results likely indicate that students reduced the list of possible correct answers from the full set of answer choices (i.e., they eliminated poor distractors), but low-ability students still struggled to identify the correct answer from that reduced set of answers.

For example, Q12 had a guessing parameter that was considerably higher than the other items. Table 8 includes the answer distributions of lower-ability students (aDCI total score of three or lower) for all items. The answer distribution for Q12 suggests that answers A, C, and (to a lesser extent) E were poor distractors. The probability of randomly selecting the correct answer out of the two remaining choices is 50% which is close to the guessing parameter for Q12. Therefore, we are not overly concerned with the high guessing parameter of Q12 because the lack of effective distractors likely explains its high value.

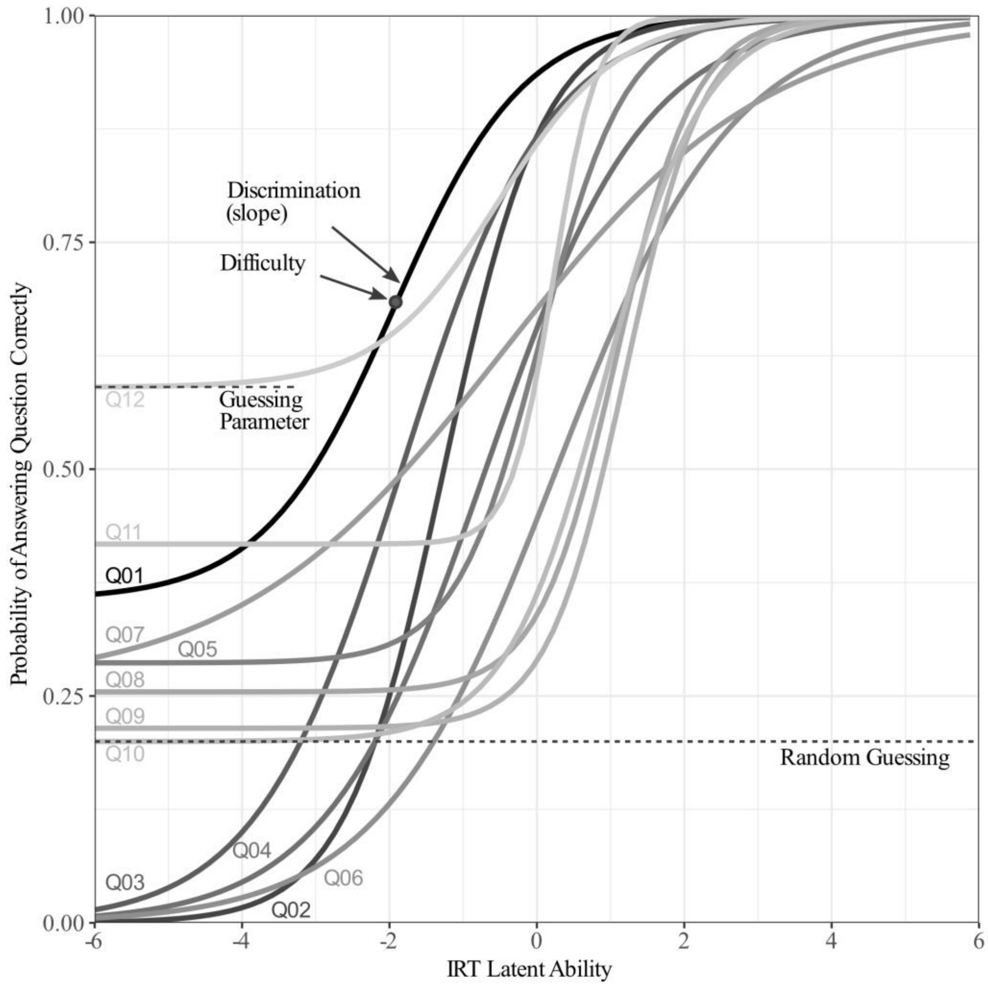

Figure 2. A 3PL IRT model was used to fit each aDCI item to determine the discrimination, difficulty, and guessing parameters. The difficulty (inflection point of the curve and the latent ability level that bisects the sample) and discrimination (slope at the inflection point) are indicated for Q01, and the guessing parameter is illustrated for Q12. The random-guessing probability is 0.2 because all items have five answer choices.

Table 8. Summary of the IRT parameters and the answer distributions for lower-ability students who answered three or fewer questions correctly on the aDCI. Correct answers are bolded.

	IRT (3PL Model)			Answer Distribution for Lower-Ability Students						
Question	Discrimination	Difficulty	Guessing	A	B	C	D	E	Multiple Selected	None Selected
Q1	1.13	-1.94	0.36	15	1	2	1	**33**	0	0
Q2	1.49	-1.26	0.00	25	**8**	17	1	1	0	0
Q3	1.02	-1.84	0.00	6	**16**	0	20	9	0	1
Q4	0.92	-0.69	0.00	17	7	1	20	**6**	0	1
Q5	1.72	0.05	0.29	25	6	**8**	11	2	0	0
Q6	0.83	0.27	0.00	24	16	**4**	7	1	0	0
Q7	0.55	-0.46	0.26	14	**13**	24	0	1	0	0
Q8	1.89	1.09	0.25	5	23	**7**	15	2	0	0
Q9	1.87	1.20	0.21	1	20	**1**	0	29	1	0
Q10	1.48	0.92	0.20	13	10	15	6	**8**	0	0
Q11	3.43	0.22	0.42	0	**7**	30	14	1	0	0
Q12	1.24	-0.52	0.59	2	15	3	**23**	8	0	1

Overall, these three conclusions from the IRT analysis support Hypothesis 2, except for Q7, in that the aDCI items have appropriate difficulty and discrimination for the students in our sample.

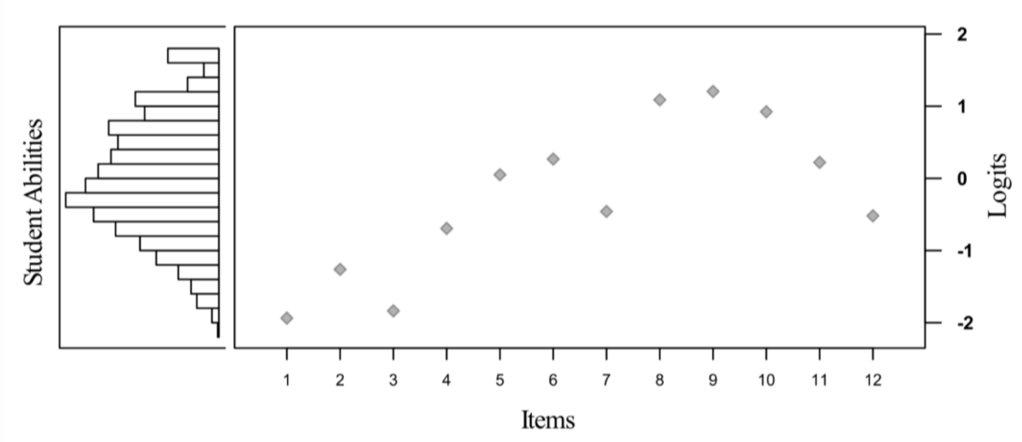

Figure 3. An item-person map (or Wright map) of the ability scores of the participants and the difficulty values for the 12 items of the aDCI. A logit (vertical axis) is the natural log transformation of 1) the odds ratio for answering an item correctly, or 2) the ratio of a student's ability divided by one minus their ability (Wilson, 2005). An item for which a student, in general, would have a 50% chance of answering correctly would have a logit of zero, and an average-performing student with an ability of zero would have a logit of zero.

Hypothesis 3: Correlation with Similar Measures of Conceptual Understanding

The correlation coefficients between the students' overall performance answering conceptual questions on the three intermediate exams for the dynamics course, their factor scores (from the CFA analysis), their ability scores (from the IRT analysis), and their total aDCI score are shown in Table 9. All of the correlation coefficients were statistically significant ($p < 0.050$), and their magnitudes with the instructor-written questions correspond with a medium effect size (Cohen, 1992), suggesting that they measured similar (if not the same) constructs as proposed in Hypothesis 3.

Figure 4 shows that the relationship between the students' total aDCI scores and the CFA factor scores (which were highly correlated with the IRT ability levels) was linear and highly correlated. This relationship allows for the aDCI total scores to be used as a proxy measurement of the students' overall conceptual understanding of dynamics without having to conduct a CFA or IRT analysis.

Table 9. Strong correlations (coefficients in the lower diagonal with standard errors in the upper diagonal) between different measures of the students' conceptual understanding suggest that they all may be measuring the same construct.

	Exam Questions	CFA Scores	IRT Abilities	aDCI Total Score
Exam Questions		0.03	0.03	0.03
CFA Scores	0.46		0.03	0.03
IRT Abilities	0.44	0.99		0.03
aDCI Total Score	0.46	0.99	0.97	

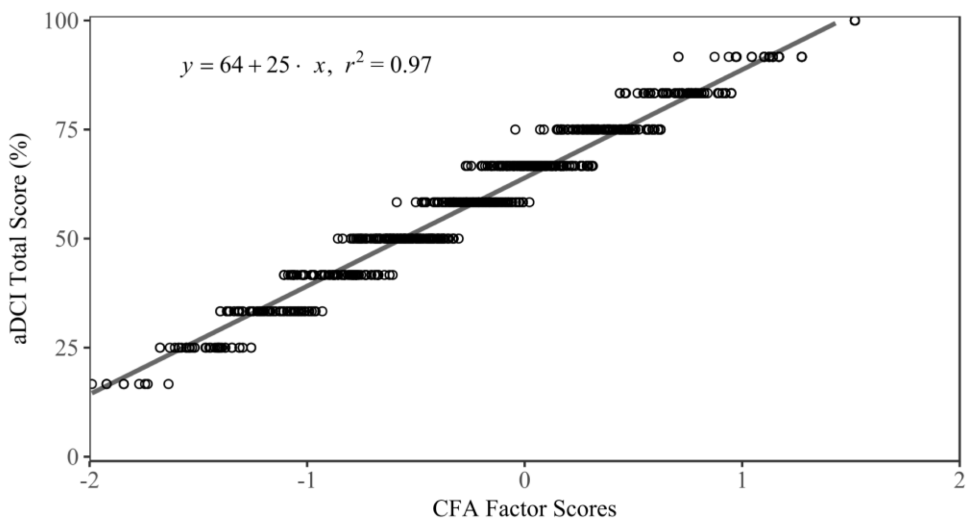

Figure 4. The students' total aDCI scores can be used as a measure of conceptual understanding because the aDCI scores are linearly related and highly correlated with the latent factor score from the CFA.

Hypothesis 4: Measurement Invariance Across Genders

The tetrachoric correlation coefficients used in the MG-CFA are shown in Table 10 (Q7 was not included in this analysis because of poor psychometrics in Hypotheses 1-3). On average, the correlation coefficients for women exceeded those for men, foreshadowing that a single-factor latent structure will fit the data from women better than it will for the aggregated or men-only data.

Table 10. The correlations coefficients between the aDCI items for men (below the diagonal) and women (above the diagonal).

Item #	Q1	Q2	Q3	Q4	Q5	Q6	Q8	Q9	Q10	Q11	Q12
Q1		**0.54**	0.27	0.20	0.03	-0.08	0.15	**0.34**	-0.08	0.18	0.13
Q2	**0.32**		**0.34**	**0.41**	**0.35**	0.26	0.30	0.27	0.16	**0.35**	0.15
Q3	0.14	**0.32**		0.27	0.14	0.27	0.08	0.16	0.10	0.21	0.30
Q4	0.16	0.24	0.29		**0.38**	0.26	0.20	**0.31**	0.03	**0.32**	0.24
Q5	0.25	**0.37**	0.29	0.24		0.12	0.26	0.24	0.11	0.11	**0.35**
Q6	0.29	0.25	**0.38**	0.19	0.24		0.10	0.24	0.14	0.26	0.12
Q8	0.18	0.15	0.05	0.26	0.24	0.12		0.22	0.03	0.26	0.20
Q9	0.22	0.21	0.02	0.19	0.19	0.09	0.28		**0.32**	**0.40**	0.18
Q10	0.19	0.27	0.06	0.15	0.28	0.22	0.20	0.18		**0.39**	0.06
Q11	0.21	**0.35**	0.17	0.26	0.24	0.22	0.25	0.29	**0.38**		0.18
Q12	0.08	0.25	0.15	0.20	0.17	0.17	0.07	0.13	0.14	0.22	

Note. The correlations coefficients greater than 0.30 are bolded. The shaded cells indicate a correlation coefficient less than 0.30 for one gender with a corresponding coefficient greater than 0.30 for the other gender. n_{men} = 1031, n_{women} = 219.

Table 7 shows the results from the MG-CFA that was used to investigate the invariance of the measurement models for men and women. The change in fit statistics for Models 1-3 indicated that the CFA models for men and women had equal form, equal factor loadings, and equal thresholds when considering all of the items in aggregate. For Model 1 (equal form), the significant χ^2 *p*-value was likely an artifact of a large sample size. For Model 2 (equal factor loadings), the positive change in CFI was likely an artifact of the CFI calculation incorporating model complexity (Bentler, 1990; Iacobucci, 2010), and Model 2 was less complex than Model 1 because of the constraints imposed on the factor loadings. Model 3 (equal thresholds—overall) had a scaled $\Delta\chi^2$ *p*-value that was much higher than the 0.050 criterion and a negative change in CFI (indicating a worse fit) that was only slightly outside of the recommended threshold of 0.01. Overall, the evidence suggested that the aDCI scores were measurement invariant at the threshold level when considering all of the items simultaneously.

In the first phase of our testing for equal thresholds for individual items, Q3 was identified as the item with the lowest $\Delta\chi^2$ *p*-value ($p = 0.032$), followed by Q6 ($p = 0.057$), Q4 ($p = 0.091$), and the rest of the items. Accordingly, Model 4 freely estimated the threshold for Q3 and was compared to Model 3. The lower χ^2 value, a $\Delta\chi^2$ *p*-value of less than 0.050, and the positive change in CFI indicated that Model 4 fit the data better than Model 3. The change in CFI of less than 0.010, however, illustrated only a small improvement in model fit. When considered together, the two model-difference statistics suggested that item Q3 was biased against women, but the magnitude of the bias was relatively small. The comparison of Model 5 (with the thresholds for Q3 and Q6 freely estimated) to Model 4 yielded a similar conclusion: Q6 was biased against women, but the bias was small.

The analysis continued in a similar fashion for Q4 (Models 6) and the rest of the items (not shown), but the change in fit indices suggested that none of these models statistically improved the goodness of fit when compared to Model 5. Thus, only Q3 and Q6 exhibited statistically-significant measurement bias across genders.

In summary, the statistical evidence suggested that the aDCI scores were measurement invariant at the threshold level when considering all of the items in aggregate. At the item level, two items, Q3 and Q6, exhibited DIF with a slight bias against women. Figure 5 summarizes the final measurement model, and Table 11 lists the thresholds for the eleven indicators.

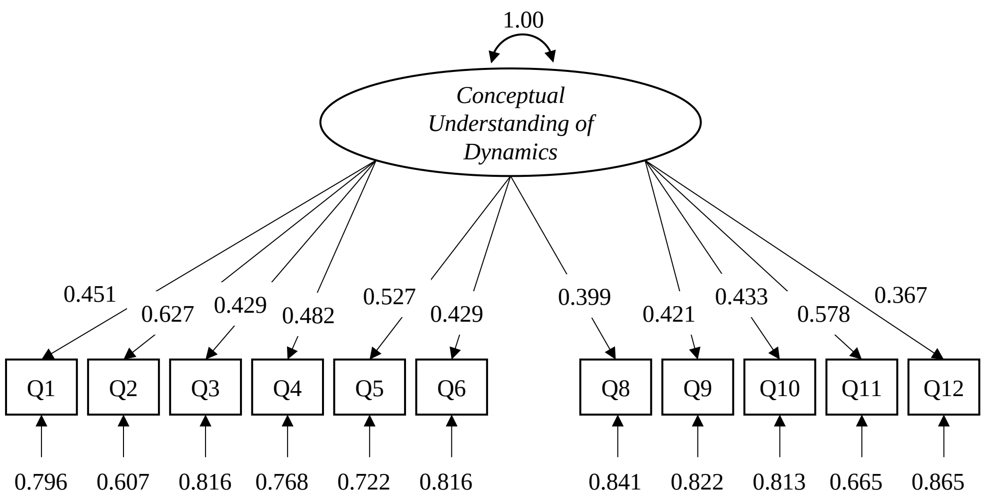

Figure 5. The test results of Hypothesis 4 showed the aDCI scores to have invariant form, factor loadings, and thresholds across men and women. All of the factor loadings were statistically significant with p-values < 0.001.

Table 11. The thresholds for nine of the eleven items were invariant across gender; Q3 and Q6 had unequal thresholds (bolded), which suggested that they were potentially biased against women. A lower threshold corresponds to a higher proportion of students in that group answering the item correctly.

	Q1	Q2	Q3	Q4	Q5	Q6	Q8	Q9	Q10	Q11	Q12
Men	-1.38	-0.89	**-1.05**	-0.39	-0.39	**0.04**	0.19	0.31	0.17	-0.48	-1.05
Women	-1.38	-0.89	**-0.76**	-0.39	-0.39	**0.29**	0.19	0.31	0.17	-0.48	-1.05

Gender Bias in the aDCI

The measurement invariance analysis suggested that Q3 and Q6 may be slightly biased against women, but the analysis does not tell why the items may favor men. To better understand the DIF of Q3 and Q6, we qualitatively evaluated Q3 and Q6 for content and context bias, as informed by our review of the physics education literature. We used Rennie and Parker's (1993) gender-orientation framework, see Table 3, to identify possible content and context effects (as defined by Ding and Caballero (2014)). We also consulted three gender-studies experts to aid in this gender-orientation analysis.

Description of the Biased Items

To maintain question security, we do not include the full copies of Q3 or Q6. Q3 originates from the FCI and involves a hockey puck sliding at a constant velocity across frictionless ice. The puck is kicked with a force perpendicular to the direction of its current motion, and the students are asked to identify the path of the puck after the kick. Five paths are pictorially provided as answer choices. Q6 describes a rear-wheel-drive car that is accelerating forward, and an image of a sports car is used to illustrate the scenario. The tires do not slip on the road, and students are asked to find the magnitude and direction of the friction force acting on the front tires. Five answer choices are provided, each with a symbolic equation for the magnitude of the friction force and a direction for the force ("to the left" or "to the right").

Bias in Q3

The categorization of Q3 according to Rennie and Parker's gender orientation framework identified the hockey context of Q3 as potentially appealing to the background experiences of men more than women. Previous FCI researchers have also advised that the hockey context of Q3 favors men (Dietz et al., 2012; McCullough, 2004). However, some of our experts suggested that the bias may be more geographical (favoring those from colder and Northern climates) than gender related. Overall, we concluded that the hockey context is a *possible* source of bias, but not a *definitive* source of bias for Q3.

Q3 of the aDCI was originally copied from the FCI, #8 in version 95 (Physport, n.d.), and research on the gender fairness of the FCI has not identified Q3 as having statistically-significant gender bias (Dietz et al., 2012; Osborn Popp et al., 2001; Traxler et al., 2018). While it is true that some of the FCI studies do not have comparable samples to the engineering students of this study, at least two of the FCI studies (Traxler et al. (2018) and McCullough (2011)) have samples from university-level, calculus-based physics courses that typically enroll science and engineering majors. We would expect students in these calculus-based physics classes to perform similarly on #8 of the FCI as the engineers in our sample do on Q3 of the aDCI.

Our results regarding the DIF of Q3 could differ from FCI research because of different analysis methods. We utilized MG-CFA to investigate measurement invariance across gender, but Traxler and colleagues (2018; who have published one of the most complete studies of gender bias in the FCI) used the Mantel-Haenszel and the Lord's statistic (an IRT-based method). Because our results suggested that Q3 (and Q6) were only slightly biased against women, the differences in samples and methods could explain the contradictions between our results and those published for the FCI.

Bias in Q6

Q6 was categorized by some of our experts as appealing to the background experiences stereotypically more common for men than women because it may be more likely that men understand the meaning of "rear-wheel-drive," a phrase used to indicate that the rear tires provide the traction force required to accelerate the car forward. This opinion aligns with Ding and Caballero's (2014) content effect because it reflects the perspective that boys in the USA are often socialized to know more about how automobiles work than girls (D. G. Johnson, 2010). However, some of our experts argued that this generalization about girls' relative knowledge of automobiles may not hold true for women in engineering because women in engineering have self-selected into a masculine-oriented field that centers on understanding how systems and machines work. Even though there was disagreement regarding the gender bias related to the phrase "rear-wheel-drive," all of the experts agreed that the sports-car image used in Q6 would be stereotypically associated with men more than women. This image could be contributing to a context effect.

Q11 and Q12 also involved a rear-wheel-drive sports car, but these two questions did not display DIF. One explanation for this difference is that Q6 requires students to understand how a rear-wheel-drive car works, but Q11 and Q12 do not require this specialized knowledge. Q11 and Q12 pertain to the kinematics of a wheel that rolls without slipping, and the question prompt only uses a rear-wheel-drive car as the structure to which the wheel is attached. Furthermore, the primary image of Q11 and Q12 is that of a wheel and tire, not the sports car. Therefore, the likelihood of the sports-car image causing a context effect favoring men in Q11 and Q12 may have been less than that for Q6 because the students' focus was on the wheel (a gender-neutral image) and not the sports car.

Discussion

Review of Purpose and Results

The purpose of this study was to evaluate the extent to which aDCI scores can be used as a reliable, fair, and valid measure of undergraduate students' overall conceptual understanding of dynamics. We organized our inquiry around four hypotheses that focused on the evidence of the aDCI's latent structure, difficulty, discrimination, correlation with similar measures, and gender fairness (in terms of measurement bias). We review the evidence for each hypothesis below.

For Hypothesis 1, the results of the CFA suggested that a single-factor latent model fit the aDCI scores well. This unidimensional latent structure reflects the intentionality of the aDCI developers to select items from the DCI that assessed a broad range of topics to approximate the students' overall understanding of dynamics. The fit of the IRT model further supports the unidimensionality of the aDCI. The correlation of the students' factor scores with their performance on instructor-written conceptual questions provides evidence for the argument that the single latent factor (of the CFA or IRT models) represents the students' overall conceptual understanding of dynamics. Therefore, the evidence for Hypothesis 1 suggests that the students' total aDCI score can be interpreted as a measure of their overall conceptual understanding of dynamics.

The results of the IRT analysis used to test Hypothesis 2 indicated that most of the items on the aDCI have appropriate difficulty and discrimination values for the sample tested. Two items were identified as especially easy (they had high difficulty values), but high performance on these items was expected because the items targeted fundamental, particle-mechanics knowledge that the students likely learned in a prerequisite physics class. All of the items, except Q7, had high discrimination values (maximum slope steepness), meaning they reliably differentiated the higher-performing students from the lower-performing students.

In addition to a low discrimination value, Q7 also had low correlations with other items and low factor loading in the CFA, indicating it might be measuring a different construct than the other items on the aDCI. Q7 was one of the DCI items Jorion et al. (2015) identified as having poor psychometric characteristics. Upon closer inspection, we found the wording of Q7 to be imprecise with multiple correct answers, depending on how the question was interpreted. Thus, multiple pieces of evidence suggest that the modification or replacement of Q7 could improve the utility of the aDCI, and a clarification of the question wording so that only one answer is correct may be all that is needed.

The results of testing Hypothesis 3 indicated that the aDCI total scores positively correlated with the students' performance on similar, instructor-written questions. The relationship between the aDCI total scores and the factor scores (from the CFA which were highly correlated to the IRT ability levels) was linear and had a high coefficient of determination. These two results provide evidence in support of the aDCI scores measuring one latent factor, and the latent factor can be interpreted as the students' overall conceptual understanding of dynamics. Wang and Bao (2010) made a similar conclusion regarding their students' conceptual understanding of physics based on the linear relationship between the students' FCI scores and their IRT abilities.

For Hypothesis 4, the analysis of measurement invariance found the aDCI to have equal form, equal factor loadings, and equal thresholds for men and women when considering all of the items in aggregate. These results suggest that, on average, the aDCI functions similarly for men and women in measuring the students' overall conceptual understanding of dynamics. When considering all items in aggregate, the aDCI scores of men and women display: the same single-factor latent structure, the same relationships between the items and the latent factor, and the same probabilities of answering a given question correctly. However, at the item level, the analysis identified two items, Q3 and Q6, that exhibited slight bias against women. The bias of these items indicates that when considering a man and a woman with equal overall understanding of dynamics, the man has a higher likelihood of answering Q3 and Q6 correctly than the woman. To understand why these items may favor men, we evaluated them for content and context bias.

The supporting and contradicting evidence for the possible sources of gender bias in Q3 and Q6 make it difficult to definitively say why these two items favor men. For Q3, the hockey context may disadvantage women. For Q6, the need to know how a rear-wheel-drive car works and/or the image of a sports car may differentially affect students' performance based on gender. The uncertainty in the sources of bias supports the need for further validation and fairness studies of the aDCI, DCI, and concept inventories in engineering more broadly.

Fairness Implications

The investigation of DIF identified two items that favored men, but this bias was not evident in the psychometric models that used aggregated data. Based on the lack of research regarding the fairness of engineering education assessments (Kerrie A. Douglas et al., 2016), it is highly likely that many researchers would have found the psychometric evidence (from Hypotheses 1-3) satisfactory for their use of the aDCI scores as measures of the students' overall conceptual understanding. However, our results suggest that instructors and researchers must consider the gender bias in at least two of the aDCI items (Q7 was not tested for DIF) when interpreting women's total scores. Two additional incorrect responses (corresponding to the two biased items) yields an almost 17 percentage-point reduction in a student's total aDCI score. While our results indicate that the bias of Q3 and Q6 is small, it undoubtedly contributes to the gender gap in the aDCI scores that has been previously reported (Prebel et al., 2017). Thus, decisions made based on a student's overall aDCI performance, including the assignment of points toward their grade in a course, unfairly disadvantage women.

Limitations and Future Work

One limitation of this study is that it was conducted with students from a single institution. As Madsen et al. (2013) determined, many findings from research on the gender gaps of physics CIs are not consistent across studies; thus, future research should consider how the aDCI functions at other institutions. Future work should also incorporate fairness studies for using the aDCI across other subgroups of students, including race/ethnicity, social economic status, academic major, and international status. Content and context effects could be especially relevant to English-as-a-second-language learners (Kerrie A. Douglas et al., 2018) and to students who have lived in cultures different than those present in the USA. An analysis of fairness for some of these subgroups (e.g., subgroups based on race/ethnicity) would especially benefit from more data because of their small sample sizes in engineering.

A second limitation of this study is the small number of women in the sample compared to men. This unbalanced sample could be hiding DIF that the measurement invariance analysis cannot detect because the fit of the model for the men's covariance matrix may have overshadowed a lack of fit for the women's covariance matrix. The sample sizes of men and women could be made equal by randomly subsampling from the pool of men, but the statistical power to detect DIF across the groups greatly decreases with this technique because of the small number of women in the sample. Given that women students are a small fraction of the overall student population in engineering, defining new norms for fair statistical models while maintaining sufficient power is a challenge for the engineering education research community.

A qualitative study of how women in the course experience the gender bias of the aDCI, as illustrated here, or other course assignments could help contextualize our findings. For example, do students (women or men) recognize gender bias in the course materials (including assessments), and does the content or context of these materials cause students to feel disadvantaged or uncomfortable? If so, in what ways do students articulate this discomfort, and what suggestions do they have for addressing it? A qualitative study, potentially including interviews with both women and men, would inform our understanding of students' experiences and could inspire changes to course materials to make them more inclusive and fair.

Conclusion

This study investigates the extent to which a student's aDCI total score can be interpreted as a reliable, valid, and fair measure of their overall conceptual understanding of dynamics. To our knowledge, this study is the first to implement an argument-based approach for the validation study of a CI (as proposed by Kerrie A. Douglas & Pellegrino, 2017) and the first to investigate the fairness of an engineering CI. The results of our study suggest that aDCI scores, excluding Q7 which should be modified or replaced, for the men in our sample can be interpreted as measures of the students' overall conceptual understanding in dynamics with evidence of: 1) broad content coverage and instructional relevancy, 2) appropriate interpretation of scores with regard to their underlying, single-factor latent construct, 3) appropriately difficult items that discriminate students based on their level of conceptual understanding, and 4) strong correlations between aDCI total scores and other measures of dynamics conceptual understanding. The total aDCI scores for women, however, incorporate two items with slight gender biases against women and, therefore, do not accurately reflect women's overall conceptual understanding of dynamics.

Unless further research refutes our results and supports the aDCI as a fair assessment tool for *all* students, or until the aDCI is modified to be gender inclusive and fair, we do not support its use in high-stakes testing, including its use on a final exam (as was done for our sample). Instead, we suggest that the aDCI, in its current form, be used as a low-stakes assessment instrument for measuring students' overall conceptual understanding of dynamics, and instructors and researchers should account for the DIF of Q3 and Q6 and the validity concerns of Q7 when making inferences from the aDCI scores. Alternatively, instructors could administer a shortened aDCI that excludes Q3, Q6, and Q7, knowing that the number of concepts assessed by a shortened aDCI would be less than the 12-item aDCI.

This work highlights the importance of designing inclusive assessments and validating their use with psychometric models that do not unfairly disadvantage certain subgroups of students—such as women. When assessments utilize validation studies that are dominated by one group of students, such as men, it is often unknown whether group differences in scores are artifacts of the assessment questions themselves, or truly representative of differences in the learners' understanding. Without evidence that the assessments themselves are truly fair for all engineering students, there is a very strong risk of educational inequity. Thus, more fairness studies of engineering education assessments are needed to better inform the academic community on what factors should be considered when designing an assessment that does not unfairly disadvantage students based on their background or socialization.

Acknowledgements

The authors would like to acknowledge the many students, staff, and faculty that made this work possible, including Craig Zywicki, Taylor Prebel, Charles M. Krousgrill, and David B. Nelson. Additionally, we would like to recognize Amy C. Moors, Cheryl Cooky, and Sharra Vostral for their expert insights on the possible sources of gender bias in the aDCI questions. The material detailed in the present manuscript is based upon work supported by the National Science Foundation under Grant No. DUE-1525671. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 3. A CLUSTER-BASED APPROACH TO IDENTIFYING AND UNDERSTANDING STUDENTS' ARCHETYPICAL RESOURCE-USAGE PATTERNS

Stites, N. A., Berger, E., DeBoer, J., & Rhoads, J. F. (in review). A cluster-based approach to identifying and understanding students' archetypical resource-usage patterns in an active, blended, and collaborative learning environment.

Abstract

Engineering educators continue to develop and implement pedagogical innovations. However, data on how these innovations impact the students' experiences and achievement is often lacking. The goal of this study was to identify and understand how and why students engaged with the resources available in an active, blended, and collaborative learning environment. We collected survey data from 581 engineering students on how frequently they used nine different resources of an undergraduate dynamics course. A cluster analysis identified nine, qualitatively-unique resource-usage patterns. We then analyzed 44 student interviews and found that students exhibited their resource-usage patterns because of personal preference, scheduling restrictions, and perceived expectations and values for a given resource. The interview data also suggested that all of the resource-usage patterns supported the students' desires to develop a deep understanding of the course material. The findings of this study provide instructors with data-driven information on the archetypical resource-usage and help-seeking behaviors of their students. Instructors can use this information to better coach their students and to design curricula and resources that support many different subgroups of students, not just the stereotypical or average student.

Keywords: cluster analysis, help seeking behavior, thematic analysis, engineering education

Introduction

Prominent reports have called for the adoption of engineering education innovations (Jamieson & Lohmann, 2012; Litzinger, Lattuca, Hadgraft, & Newstetter, 2011), and active, blended, and collaborative learning are pedagogical strategies that continue to increase in popularity. Active learning involves physical activity (Chi, 2009), blended learning combines in-class instruction with online learning outside of the classroom (Bernard et al., 2014), and collaborative learning incorporates students working in groups to attain a shared goal (D. W. Johnson et al., 1991). Studies have shown that active, blended, and collaborative pedagogies improve student learning (Bernard et al., 2014; Freeman et al., 2014; Stump, Hilpert, Husman, Chung, & Kim, 2011). However, in learning environments that enable the self-regulated use of blended resources for active and collaborative learning, it is often unclear how, and to what extent, students engage with the myriad resources available to them. Without having strong evidence for how students engage with the resources, instructors and course designers rely on assumptions, perceptions, and stereotypes for designing and improving educational resources and learning environments (Minichiello & Jouffray, 2018; Turns, Borgford-Parnell, & Ferro, 2015). This work aims to better understand how and why undergraduate engineering students engage with a variety of resources within a learning environment specifically designed to encourage active, blended, and collaborative learning.

According to Makara and Karabenick's (2013) proposed expectancy-value model for resource selection, students choose their help sources based on four main factors: 1) the perceived availability of the source, 2) the perceived likelihood that the source will provide help if asked, 3) the alignment between the type of help provided by the source and the type of help desired, and 4) the perceived quality of the help from the source. Regarding the type of help desired in the third factor, researchers usually discuss two types of help-seeking behaviors (HSBs): adaptive and expedient (Er & Orey, 2017). A student's HSB is considered adaptive—also referred to as strategic or instrumental, but instrumental HSB should not be confused with instrumental motivation (Gardner & Lambert, 1972)—when the goal of the action is to understand the material and to become a more-autonomous learner in the future (Horowitz, Rabin, & Brodale, 2013; Karabenick, 2003; Karabenick & Berger, 2013; Newman, 2002). Conversely, expedient HSBs (also known as non-adaptive or executive) are characterized by a student seeking nonspecific help—e.g., when a student asks for help before they even try the problem—or help that leads to the correct completion of the task with as little effort as possible, which perpetuates their dependency on others to solve problems (Er & Orey, 2017; Karabenick, 2003). The expectancy-value model proposes that students with shared help-seeking goals and similar perceptions of the expectations and values for the available help sources may choose to utilize similar help resources.

To date, researchers of academic HSBs have only considered the students' use of a given resource individually, rather than as a part of an overall resource-usage pattern (e.g., Hao, Barnes, Branch, & Wright, 2016; Horowitz et al., 2013). Multiple analyses of individual resources can provide information about the average use of a given *resource*, but they do not necessarily depict the holistic resource-usage behaviors of a given *student*. Resource-centered approaches provide little information about what *combinations* of resources were most prevalent or to what extent the students of a given resource-usage pattern shared common perceptions of the expectations and values of the resources they did or did not use. We posit that most students in technical courses, including engineering-sciences courses, utilize a combination of help sources, rather than a single resource, and research has shown this to be true for students in the course of this study (Evenhouse, Kandakatla, Berger, Rhoads, & DeBoer, in preparation; Kandakatla, Berger, Rhoads, & DeBoer, in review). Therefore, we contend that a student-centered approach is more appropriate than a resource-centered approach when trying to understand the holistic, archetypical resource-usage behaviors of multiple subgroups of students in technical, resource-rich courses.

Some researchers have recognized the value of grouping students together by common HSB characteristics (e.g., Karabenick, 2003). However, these researchers have grouped students according to their general help-seeking tendencies, which may not provide detailed enough results to inform the design or modification of a specific course's resources. Instead, this study exemplifies how researchers can group students according to their holistic usage behaviors of the specific resources that are available in a given course. Instructors can use this data-driven, resource-usage information—rather than assumptions or stereotypes—to better design and foster a learning environment that supports multiple, diverse subgroups of students (Minichiello & Jouffray, 2018; Turns et al., 2015).

Purpose of Study

This study aims to understand how and why students use the resources available in an undergraduate dynamics class (hereafter referred to as Dynamics, with a capital "D," whereas the field of dynamics will be referred to with a lower-case "d") taught within an active, blended, and collaborative learning environment, called Freeform.

The following research questions guide this work:

RQ1. When considering all of the resources of Dynamics simultaneously, what are the students' archetypical patterns of resource usage?

RQ2. How and why do students enact their respective resource-usage pattern, and to what extent do the students' perceived expectations and values for the resources influence their resource usage?

This work is part of larger project researching the students' resource-usage patterns in Dynamics and the extent to which the students' patterns explain their performance in Dynamics. We focus on identifying and understanding the students' resource-usage patterns in this paper, and we investigate how the students' resource usage relates to their performance in Dynamics in a companion paper (N. A. Stites, E. Berger, J. DeBoer, & J. F. Rhoads, in review-b). Knowledge of how and why the students use the available Dynamics resources coupled with their performance in the class could help instructors better coach students on how to be successful in the course, and it could guide the modification or development of resources to better support the students' learning (Turns et al., 2015).

To our knowledge, this study is the first research on HSBs to employ a model-based clustering technique that groups students according to their self-reported usage data. After identifying the "clusters" (groups) of students who exhibit the same resource-usage pattern, we utilize student interviews to better understand how and why the students of each cluster used the resources as they did. We view our results through a conceptual framework based on Makara and Karabenick's (2013) expectancy-value (EV) model for resource selection.

Giblin (2016) also viewed their qualitative results through a framework based on Makara and Karabenick's EV model, and they found evidence that all of the EV factors, except one, influenced the resource selection of their upper-level-mathematics interviewees. It was unclear if the factor regarding the alignment of the type of help provided and the type of help desired influenced the students' source selection because all of Giblin's interviewees described their HSBs as adaptive in nature. This study uses a similar sample, engineering students, in a similarly technical course. Therefore, we expect to find evidence in the interviews that corroborates Makara and Karabenick's EV model of resource selection, but we may not find evidence to support or refute the contention that the alignment of the help type affects the students' decisions.

Study Context

In 2010, two engineering instructors implemented a new learning environment, called Freeform, for teaching Dynamics (Rhoads et al., 2014). The Freeform learning environment was designed to align with the known benefits of active (Freeman et al., 2014), blended (Bernard et al., 2014), and collaborative (Wiggins, Eddy, Grunspan, & Crowe, 2017) learning. The Dynamics instructors are encouraged to incorporate active and collaborative pedagogies in their classrooms, and an online discussion forum provides students with a way to asynchronously ask each other questions about the materials and their assignments. Freeform also includes a custom-written textbook, called a lecturebook, in which students write their notes and their solutions to example problems. Each section of the lecturebook begins with a short theory section and ends with the problem statements of many unsolved examples that have a problem-solving or conceptual focus. An online solution video accompanies every example problem in the lecturebook (excluding the conceptual problems) and every homework problem. Finally, the learning environment leverages a tutorial room that specifically supports statics and dynamics courses, and this "help room" is staffed by student teaching assistants (TAs) about 8-10 hours a day, six days a week. The distributed hours of the help room are essential because the Dynamics students have two homework problems due three times per week. Dynamics also has three intermediate exams and one final exam. The integrated suite of Dynamics resources is designed to accommodate a variety of help-seeking preferences as students prepare for, or complete, their assessments.

The general perceptions from Dynamics instructors and a limited amount of aggregated, course-level data suggest that students utilize certain Dynamics resources frequently (e.g., online videos) and others hardly at all (e.g., instructors' office hours) (Evenhouse et al., in preparation;

Rhoads et al., 2014). However, the Dynamics resources are purposely designed to be aligned and integrated with one another, and the holistic patterns of resource usage that the students exhibit is unknown. Using a combination of cluster analysis, thematic analysis, and a conceptual framework based on an expectancy-value model for resource usage, this study seeks to understand the Dynamics students' resource-usage patterns to inform improvements to the Dynamics resources and to guide instructors on how to coach their students to be successful in the course (Minichiello & Jouffray, 2018; Turns et al., 2015).

Conceptual Framework

Importance of Help-Seeking Behaviors

A student's help-seeking behaviors (HSBs) are commonly considered a strategy for self-regulated learning (Karabenick & Berger, 2013). Pintrich and Zusho's (2007) model of motivation and self-regulation posits that self-regulatory processes, like HSBs, can have direct and indirect effects on student's achievement. Resource-usage patterns are an outcome of the students' HSBs, so they too can relate to achievement. Thus, by better understanding their students' resource-usage behaviors, instructors further their insights into the factors that can influence their students' achievement.

Help Seeking and Self-Regulated Learning

The help-seeking process is considered a strategy for self-regulated learning (SRL) because it is a cyclical and reflective process in which students continually modify their behavior to better support their goal attainment (Herring & Walther, 2016; Horowitz et al., 2013; Newman, 2002). The help-seeking process has been modeled as having eight components, see Figure 6, and those eight components align with the three phases of Zimmerman's model for SRL (Karabenick & Berger, 2013). This study focuses on the resource-usage outcomes of soliciting and obtaining help (Steps 6 and 7) and why students used the resources they did (Steps 4 and 5).

Because SRL is a cyclic and reflective process, a student's constant evaluation of the assistance they receive from specific resources (Step 8) could lead students to exhibit different resource-usage patterns depending on the type of help they desire (Step 4). A student with a mastery-goal orientation, reflecting the student's preference for understanding rather than simply getting the correct answer, could settle on the use of a different set of resources than a student who is performance-goal oriented and considers learning as secondary to completing the task correctly and quickly. However, there are other factors that influence a student's decision of what help source to consult, as elucidated by Makara and Karabenick's (2013) expectancy-value model for resource selection.

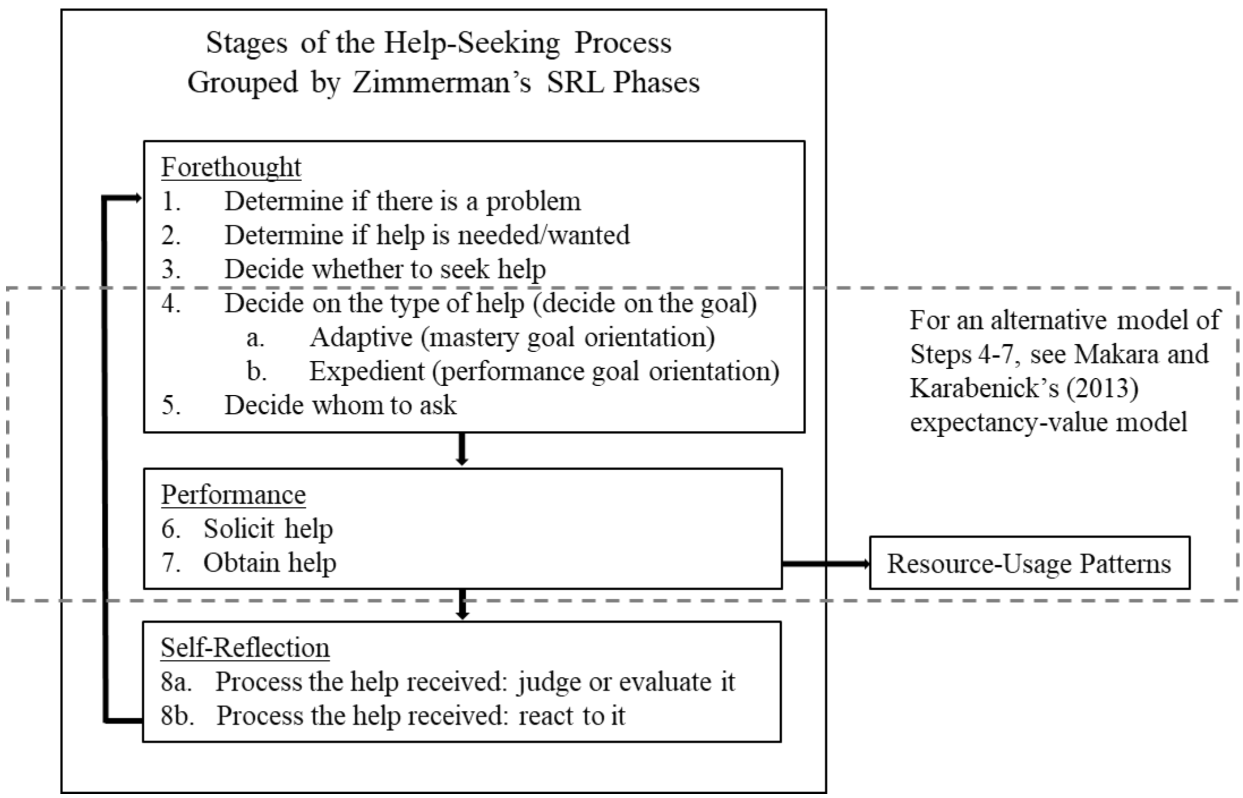

Figure 6. A self-regulated-learning perspective of HSBs underpins the conceptual framework of this study (Karabenick & Berger, 2013).

Expectancy-Value Model for Help Source Selection

Makara and Karabenick (2013) proposed an alternative model for Steps 4-7 of the help-seeking process that is based on expectancy-value (EV) theory. The EV model, shown in Figure 7, posits that the interaction of a student's expectations of a help source and the student's perceived value of the help source affect whether or not a student will seek help from that source.

Makara and Karabenick proposed that the expectancy component of this model consisted of the student's perceptions of the accessibility and availability of the source and the student's perceived expectations regarding whether or not that the source will provide help if asked. Availability refers to whether a source could provide help, and accessibility refers to how easy that help can be obtained. The value aspect of this EV model clearly articulates that the type of help desired (adaptive or expedient) and the perceived quality and accuracy of the help directly influence which help source is chosen.

Overall, the EV model suggests that if a student perceives a source as being available and willing to help *and* the type of help the source provides is perceived as accurate and in alignment with the type of help they desire, then the likelihood of a student seeking help from the is source is high. Conversely, if the source is perceived to be unavailable, unwilling to help, misaligned with the type of help desired, or inaccurate, then the likelihood of the student seeking help from that source decreases. This EV model provides a framework through which we can view our quantitative and qualitative results to better understand which expectation and value factors contributed to the resource-usage patterns of specific subgroups of students.

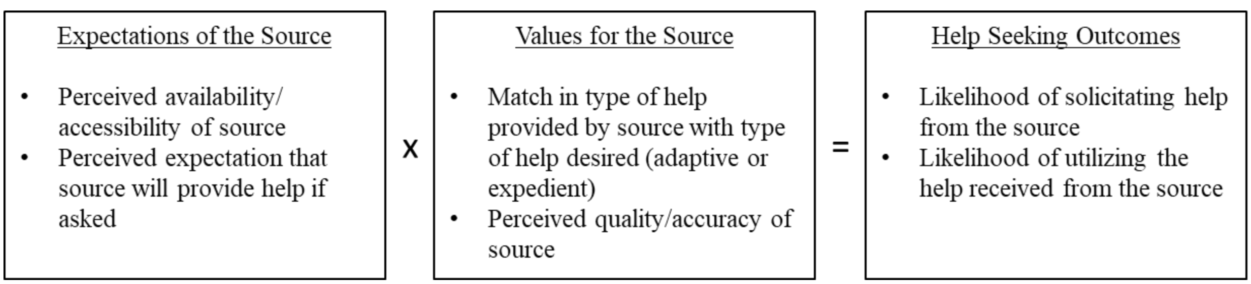

Figure 7. An expectancy-value model for help source selection (Makara & Karabenick, 2013).

Prior Dynamics Results and the EV Framework

Prior research that considered all of the students in Dynamics in aggregate, not grouped by resource-usage pattern, identified that the convenience and availability of a resource (Evenhouse et al., in preparation) and the alignment of a specific resource with the task at hand (Kandakatla et al., in review) were primary determinants of what resources students used. These findings, when viewed through the EV framework, suggest that the perceived accessibility and availably of a resource along with its perceived quality had a significant impact on whether or not a Dynamics student used that resource. However, neither of these studies mentioned that the alignment of the type of help provided by a source (adaptive or expedient) and the type of help desired by the student factored into a student's resource-selection decisions. Overall, the previous research on the HSBs of Dynamics students contributed to our earlier hypothesis that we will find evidence that the factors in Makara and Karabenick's EV model influenced the Dynamics students' resource-usage decisions, with the possible exception of the help-alignment factor because, like Giblin (2016), we may find that all of the students self-report their HSBs as adaptive in nature.

Methods

This study employed an embedded research design (Creswell, 2012, Chapter 16) because quantitative and qualitative data that related to the same phenomenon, resource usage and HSBs, were collected from the students simultaneously, during the same semester. Furthermore, neither data source directly influenced the collection of the other, and the two data sources were collected to answer different research questions in this study. Quantitative survey data were used to identify clusters of students who exhibited the same resource-usage patterns (RQ1), and qualitative interviews provided insights into how and why students used the resources in certain ways (RQ2).

Participants

The data for this study were collected from students enrolled in Dynamics, a sophomore-level engineering course at a large, public university in the Midwestern USA with the highest category of research activity (Indiana University Center for Postsecondary Research, n.d.). The sampling frame for this research was all of the students enrolled in the fall or spring semesters of Dynamics from Spring 2016-Spring 2018. Of the 1,379 students in the sampling frame, 581 completed the survey that provided the quantitative data to investigate RQ1, and the interview transcripts of 44 students served as the qualitative data for RQ2. Additional details regarding our recruitment and sampling processes for the survey and interview data sources are included in Appendix G.

The demographic characteristics of all of the participants are shown in Table 12. This demographic data was obtained from the institution's Registrar, and the categories used in Table 12 reflect how the data were collected. Gender was reported by the institution as a binary variable, and we acknowledge that this is a simplification of the gender spectrum and that the terms "male" and "female" are terms to describe one's sex, not gender. We also recognize that race, ethnicity, and international status were all confounded together into one "ethnicity" variable. Nonetheless, the demographic characteristics in Table 12 help us better understand the backgrounds and socializations of the students in our sample.

Table 12. Demographic characteristics of the survey and interview participants.

	Survey Participants		Interview Participants	
Variable	Count	%	Count	%
Major				
Mechanical Engineering	459	79%	27	61%
Agricultural Engineering	32	6%	4	9%
Nuclear Engineering	28	5%	2	5%
Multidisciplinary Engineering	25	4%	5	11%
Other	37	6%	6	14%
Ethnicity (Race/Ethnicity/ International Status)				
Domestic, White	379	65%	30	68%
Domestic, Asian	31	5%	3	7%
Domestic, URM	25	4%	0	0%
Domestic, Other	34	6%	1	2%
International	112	19%	10	23%
Gender				
Male	437	75%	28	64%
Female	144	25%	16	36%

Note. The sum of the percentages for the ethnicity categories for the survey participants does not equal 100% because of numerical rounding.

Data Sources

At the end of the semester, students were asked to complete a survey about their study habits, help-seeking behaviors, resource usage, and general experiences in the class. This study utilized one specific multiple-part question from the survey that asked the students to: "Please identify how frequently you use each of the following resources for help in Dynamics." The response options were (verbatim, and in the order in which they appear on the survey): at least once per day, 3-6 times per week, 1-2 times per week, 1-3 times per month, 1-3 times per semester, and never. The nine resources included in the survey question are listed in Table 13 along with their descriptions and their median response.

Table 13. A description of the nine resources included on the end-of-semester survey and the median frequency with which students used the resource (N = 581).

Resource	**Description**	**Median Frequency**
My peers in the class	Group quizzes in class; virtual or in-person collaboration outside of class	1-2x/wk
The course lecturebook	Combination of a workbook and concise textbook; students write notes and solve problems directly in book	3-6x/wk
The lecture example and homework solution videos	Screencasts of the instructor solving a problem; every lecturebook example and homework problem has a solution video	1-2x/wk
The course blog	"Blog" most often refers to the discussion forum, but could also be interpreted as the course website	1-2x/wk
The instructor, by asking questions in class	Could include questions before, during, or after class	1-3x/semester
The instructor, during office hours	Office hours were usually 1 hour long, 2-3 days/wk	Never
Online resources not accessed from the course blog (ex: online lectures or videos not associated with the course)	Could include online videos, online example problems, or online tutoring websites	1-3x/semester
Other students I know who are not currently enrolled in the class	Friends who have taken Dynamics previously (although there is evidence in the student interviews that students may have misinterpreted this as asking about students in other sections of the course)	Never
The TAs in the Mechanics Tutorial Room	A dedicated help room staffed over 40 hours/wk with undergraduate- and graduate-student TAs	1-3x/semester

For RQ2, the students' resource-usage behaviors were explored through semi-structured interviews conducted with students during the last week of the semester in which they were enrolled in Dynamics. A predetermined set of questions probed a variety of topics including (but not limited to): the student's perceptions about the learning climate at the institution and in their major department, their preferred study strategies, their perceptions of the quantity and quality of the resources for Dynamics as compared to the resources provided for other engineering courses, their resource usage, and their recommendations to future students on how to be successful in Dynamics. The same set of questions was used for every interview, but the interviewer could reorder the questions and/or ask follow-up and clarification questions, as appropriate, based on the interviewee's responses. Because the interviews were limited to approximately 30-45 minutes yet were used to collect data for this study and several others, not all of the resources listed on the survey were explicitly discussed in the interview. For example, students were not directly asked about their use of other online resources not provided by the instructor or about their use of students not currently in Dynamics. The audio of each interview was recorded and subsequently transcribed by a third-party transcription service.

Data Analysis

Quantitative Analysis

To quantitatively find the students' archetypical patterns of resource usage across all nine of the resources listed in Table 13, we conducted a model-based cluster analysis using the *mclust* package (version 5.3) in R (version 3.3.2) to evaluate 14 different clustering shapes with the number of clusters ranging from one to ten. The frequency-of-use data from the survey for the nine resources were on the same ordinal scale, so no data transformations were needed. Two of the primary advantages of model-based clustering over the commonly-used K-means clustering technique are that multiple shapes for the clusters are considered and the clusters can overlap because the classification of a student into a cluster is based on a vector of probabilities corresponding to the alignment of a student's behavior with that of the other students' behaviors in that cluster (Fraley & Raftery, 2007). In contrast, K-means clustering separates students into exclusive groups that are spherically or circularly shaped (Jain, 2010). Because we had no *a priori* knowledge about the shape of the resource-usage clusters and because we planned to use each student's vector of probabilities to gauge how well their behavior aligned with each cluster's typical behavior, this study utilized model-based cluster analysis.

The selection of the best-fitting cluster model was primarily driven by the Bayesian information criterion (BIC), which is a likelihood criterion that penalizes models with increased complexity (Spiegelhalter, Best, Carlin, & van der Linde, 2014). The shape and cluster-number combinations that had the three highest BIC values were considered the best-fitting models. The differences in the BIC values between the top three cluster models were small (less than 0.5% of a difference in the BIC values), so we compared the three models for differences in the number of qualitatively-unique patterns of resource usage. The most parsimonious model that also captured all of the qualitatively-unique usage patterns was selected as the final model.

Qualitative Analysis

To better understand why students utilized the resources in certain ways, we conducted what Merriam (2009) referred to as a basic qualitative study with the data from the student interviews. The focus of this basic qualitative study was to better understand the usage behavior that made each cluster qualitatively unique. The interviews of students within a given cluster were evaluated for common themes regarding why students in that cluster used a resource differently than many of their peers. We used a thematic-analysis process based on the guidelines of Braun and Clarke (2006), as described in Appendix G. After the themes were developed, we viewed our results through the EV framework to determine which expectations or values for the resources were influencing the students' decisions on which resources they used.

Only interviewees with a cluster-classification uncertainty of less than 0.30 were included in the qualitative analysis because we wanted to understand the *archetypical* behaviors identified by the cluster analysis. A student's cluster-classification uncertainty is calculated as unity minus the maximum probability in the vector of cluster probabilities. The process of determining the threshold of 0.30 is further explained in Appendix G. Because this was an embedded research design where the survey and interview data were collected simultaneously, but independently, the number of interviewees in each cluster varied from one to 12, with the majority of clusters having 3 to 6 interviewees (see Figure 8 for the distribution of interviewees across the clusters). Given the small number of interviews per cluster, we considered similar content across interviews to be a theme when it was mentioned by two or more students (for the clusters with more than one interviewee).

Results

Cluster Analysis of Survey Data

Model Selection

The cluster model that had the highest BIC (-16,490) had eight clusters, and the models with the second- and third-highest BIC values (-16,533 and -16,549) had nine and ten clusters, respectively. The model with nine clusters suggested that a group of students infrequently used the online blog (which, according to our interviews, students most-often interpret as the discussion forum, but they could have also interpreted the "blog" to be the website as a whole), a usage behavior that was not reflected in the eight-cluster model. The ten-cluster model did not offer any additional, qualitatively-different usage behaviors compared to the nine-cluster model; the additional cluster was made up of students who asked the instructor slightly fewer questions and who used other online resources outside of those provided by the instructors a little more frequently than the students in an existing cluster of both the eight- and nine-cluster models. Given the ordinal scale on which students indicated their frequency of use of these resources, we did not believe the extra cluster of the ten-cluster model provided any more information about the students' resource-usage patterns than the nine-cluster model. Thus, the nine-cluster model was chosen as the most parsimonious model that still captured the qualitatively unique resource-usage patterns of the students. The average values for how frequently the students within each cluster used each of the nine resources of the survey are shown in Figure 8.

In the nine-cluster model, the only two clusters that were substantially similar, C_2 and C_6, primarily used the same four resources, but students in C_2 used the other five resources slightly more often than the students in C_6. While one could argue that C_2 and C_6 did not exhibit qualitatively different resource-usage patterns, these two clusters were identified in models with seven, eight, and ten clusters also, so changing the number of clusters did not resolve the issue of having two similar clusters. We considered combining the similar clusters but decided to keep them separate in case the qualitative analysis yielded distinct differences between the clusters.

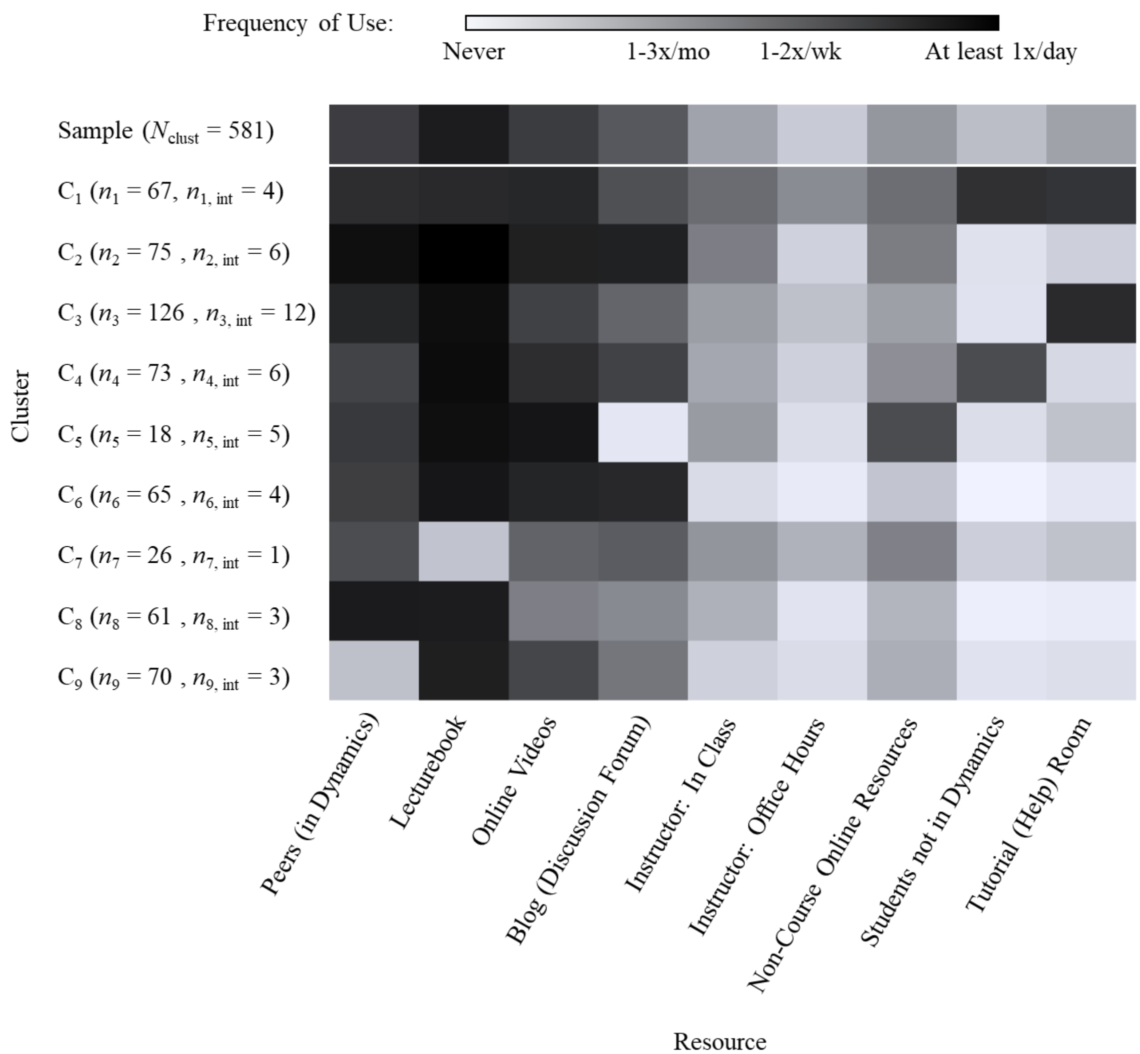

Figure 8. The average values for how frequently the students within a cluster used each of the nine resources. The sample sizes with the "int" subscripts indicate the number of interviewees in each cluster.

One measure of how well the cluster model fits the data is the uncertainty associated with the cluster classification of each student. For the nine-cluster model, almost half of the students had an uncertainty of less than 2%, and approximately 84% of the students had an uncertainty of less than 30% (which was the filtering threshold used for the qualitative study). These uncertainties were similar to those for the eight- and ten-cluster models, which had 85% and 83% of the students, respectively, with less than 30% uncertainty. The mean and median uncertainty for the nine-cluster model was 11% and 2%, respectively, which leads us to have high confidence in most of the students' cluster classification.

Characteristics of Resource-Usage Patterns

Figure 8 illustrates that a finite number of patterns represents the resource-usage behaviors of most students. Cluster 1 (C_1), on average, utilized the Dynamics resources the most often, and students in C_9 used the resources the least. This frequency of usage across clusters correlated with the students' survey responses regarding the time they spent on the class outside of lecture. For example, the number of hours per week that the students in C_1 ($M = 10.34$, $SD = 5.23$) spent on the class was statistically higher than the number of hours spent by student in C_6 ($M = 7.8$, $Med = 7$, $SD = 3.0$; $U = 2862$, $p = 0.002$, $r_{\text{effectsize}} = 0.273$, small), C_8 ($M = 7.0$, $Med = 6$, $SD = 3.5$; $U = 2955$, $p < 0.001$, $r_{\text{effectsize}} = 0.387$, medium), and C_9 ($M = 7.2$, $Med = 6$, $SD = 3.4$; $U = 3285$, $p < 0.001$, $r_{\text{effectsize}} = 0.349$, medium), where U is the Mann Whitney U test statistic and $r_{\text{effectsize}}$ is the non-parametric, point-biserial correlation effect size that is categorized according to Cohen's suggested ranges (Cohen, 1988; Fritz, Morris, & Richler, 2012).

The students in every cluster consistently used at least two of the "core" resources of Dynamics, which we classify as peers, lecturebook, online videos, and the discussion forum. Thus, according to the EV model, these core resources must generally be perceived as available, willing to help, able to provide help that matches their desired type of help, and able to provide quality and accurate help. This logical inference aligns with the finding of Evenhouse et al. (in preparation) when they qualitatively analyzed the HSBs of a similar, but aggregated, sample of students and found that the convenience and availability of the core resources contributed to their high usage.

The most common, holistic pattern of resource usage, that of C_3, mostly included the use of core resources and the tutorial room. The tutorial room is one of the few resources that has fixed times for its availability, so, according the EV model, its inclusion in the resource-usage pattern of the largest cluster likely indicates that many students perceived the value of the help it offered to be high.

As expected, the resource-usage patterns suggest that students did not use just one resource when seeking help. With nine resources and nine clusters, it is conceivable that each of the clusters would be centered around the frequent use of a single resource. Instead, every cluster of students used multiple resources. This likely indicates an awareness of the different help sources for diverse needs. However, it also likely reflects some students' lack of SRL skills and an inability to match the help source to their needs, thereby causing them to consult multiple resources before getting the appropriate help. We revisit this issue of SRL skills in the qualitative analysis of the interview data from students in C_1 because students in C_1 use the most resources.

Almost every resource was frequently used by at least one cluster of students. The least utilized resources were those involving the instructor, which corroborates the findings of Wirtz et al. (2018), who studied the HSBs of students in mechanical engineering at the same institution as our participants. Thus, the lack of using the instructor as a help source could reflect the departmental culture rather than the course-specific culture. It could also reflect the importance of instructors actively fostering an atmosphere of help seeking—which can be especially important in large classes (Griffin et al., 2014)—that reduces the threat associated with hierarchical power relationships that some students perceive between instructors and students (Er & Orey, 2017; Joyce, 2016; Martin & Myers, 2006).

Thematic Analysis of Interview Data

In this qualitative analysis, we focused on how and why students enacted their respective resource-usage pattern and how the students' perceived expectations and values for the resources influenced their resource usage. Therefore, we analyzed the student interviews for resource-usage themes within a cluster, rather than looking across clusters for themes regarding a specific resource, which has been done previously (Evenhouse et al., in preparation). We explain the resource-usage behaviors of students in C_1 for all nine of the resources; thus, the thematic analysis of C_1 is very thorough. For the other eight clusters, we briefly discuss the resource-usage characteristics that make that cluster unique. We highlight how each cluster's themes relate to three of the four factors of the EV model for resource selection: 1) perceived availability and accessibility, 2) perceived likelihood of the source providing help, and 3) perceived quality and accuracy. The one component of the EV model that we do not discuss until the end of this section is how well the help provided by the source matches the desired type of help (adaptive or expedient). We save this discussion until the end because all of the interviewees described their HSBs as adaptive, regardless of their resource-usage pattern.

Cluster 1 (Frequent Users of Most Resources)

The students in this cluster perceived the culture in engineering as collaborative, and they did not mind reaching out to their peers, TAs, or instructor for help with Dynamics. For example, one student commented:

> [The undergraduate student culture is] good in the sense that a lot of people seem to want to help each other with understanding the concepts behind their classes. You know, I'll ask someone a question about a homework problem—and I can go up to practically anyone in the [mechanical engineering] building—and they'd be willing to help me through it. (Student 3)

Students in this cluster utilized many of the Dynamics resources when they did not understand a concept. The following quotes illustrate typical resource-usage behaviors in this cluster, all of which include some level of peer collaboration:

> I do [the homework] myself first. If I don't get it, I'll look up [an] example [from the] lecture or lecture example videos online. And then, if I still don't get it, I'll go into the help room. And there's a lot of people there, too. (Student 1)

> There're a couple guys that I knew before this…course, and we usually meet in the [mechanical engineering building]. And we'll just, like, we'll start the problems. If we don't understand it, we'll go into the help room and get our questions answered, and, obviously, along the way we'll try to help each other and that's usually how I approach the homework problems. (Student 3)

> I always review [for the] exam with my friend, but homework I always do myself—but sometimes we will discuss it a bit. (Student 4)

Multiple interviewees checked the online discussion forum for helpful hints, but they often did not find the forum that useful, in part because (in certain semesters) not many students were posting questions or answers:

> I check [the discussion forum] while I'm doing homework sometimes, just to see if there's anything interesting or anything that I might be missing from this problem. (Student 1)

> Yeah, I did not use the actual [discussion forum] this time. Like I didn't respond to homework questions or ask questions on that, but the main reason for that was because I would just go to the help room and talk to the TA about it. But yeah I felt like that wasn't as useful because not many people seem to be using it. So, if you had a question it probably wouldn't get answered on the [discussion forum]. (Student 3)

One student mentioned that they utilized their friends who had already taken Dynamics:

> I do study alone, but sometimes I have friends who already took their ME Dynamics before this semester, so I approach them and ask them questions ...[and] if they still have resources from their semesters so that I can use it and study—or just ask them for help, or something. (Student 2)

Lastly, the instructor's office hours were used infrequently, but the students found the instructors' responses to questions in class helpful. For example, Student 2 said, "Sometimes, people don't understand, and the professor will give you extra information from that. That's one thing I like about the community, … they ask questions." Overall, the students in C_1 found the resources to be very helpful, as this student succinctly articulated:

> If you're struggling, there are a lot of resources that you can go to…there are plenty of staff that you can get help from. Yeah, I felt like it was very well, I guess, instructed; well organized. (Student 1)

When considering the common behaviors of C_1 through the lens of the EV model, the students perceived many of the resources as available and willing to provide help, but their perceived quality for the resources varied. The perceived quality and the likelihood of receiving help for the discussion forum was lower because the posted discussion *might* expose misconceptions and an asked question *may* be answered. The tutorial (or "help") room was referenced by multiple students as a place to go after first seeking help from other resources; thus, the students in C_1 appear to perceive the quality and accuracy of the TA's help in the tutorial room as being higher than the other resources, with the exception being the instructor. The interview data was inconclusive on whether the use of many resources is a sign of seeking deep understanding or the result of inefficiently aligning one's needs with the help source. Because these interviewees perceived the resources as being available and willing to provide help, it is possible that they move quickly from one resource to another if they do not immediately find the help they seek.

Cluster 2 (Primarily Core Users)

Students in C_2 more frequently used the discussion forum while working on their homework when compared to C_1, possibly driven by their instructor encouraging their use of the discussion forum:

> I used [the discussion forum] to do a quick check… [at] the beginning of the semester I used it a little bit more when [our instructor] was like, "Make sure you're using this." Then when [our instructor] wasn't saying, "Use it," I would just kind of forget about it. I think it was moderately useful. I did maintain that skepticism about some of the things that were on there—that people could be wrong. It was lower on my list [of useful resources]. (Student 10)

As mentioned by Student 10 (above) and Student 6 (below), the information on the discussion forum was not always useful to the students:

> Towards the end of the semester, I checked [the discussion forum] probably every homework. … I thought it was useful just because different people brought up different things about the problems that I wouldn't think about. Or, you would get stuck and somebody else would have the same problem, and someone would give an answer. But then on the other hand, sometimes people would give answers that they wouldn't explain fully. So, it would be difficult to get what they were trying to tell you if they didn't explain it very well. (Student 6)

Thus, the students in C_2 perceived the discussion forum as more available and more likely to provide helpful insights than students in C_1, but the students in C_2 shared the same low perception of the discussion forum's quality and accuracy.

Compared to the students in C_1, the students in C_2 did not utilize the tutorial room or instructor office hours as often. Interviewees indicated that the availability and accessibility of various online resources reduced the need for students to seek help from the TAs, as exemplified by this quote:

> I've been in the tutorial rooms but less [for Dynamics] because the online stuff is more available. (Student 10)

The difference in office hours use appeared to be primarily a result of scheduling conflicts that reflected the limited availability of the instructors' office hours. However, one student referenced the power differential, an aspect of accessibility, between the instructor and student as a cause for not asking the instructor questions:

> I have class during his office hours, and I'm also like a bit intimidated by approaching him and sometimes like I don't want to ask stupid questions. So, like for me, I would try to work by myself first, and if I really cannot understand, I would just go to help room or ask my peers because I don't want to appear stupid to the instructor. (Student 5)

Overall, the students in C_2 seemed to be heavily influenced by the availability and accessibility of the resources, but quality played a role in how likely the students would be to utilize the help, especially for the discussion forum.

Cluster 3 (Core + Tutorial Room Users)

This was the resource-usage pattern with the largest membership, and compared to most of their peers, the students in C_3 frequently used the tutorial ("help") room. They preferred the immediate, personalized, and accurate help that the help room and its TAs provided to them. The help room did have only certain hours of availability, but interviewees organized their study schedules around this availability so that the help room was perceived as being highly available. For example, Students 11 and 17 took different approaches to solving their homework, but they both relied heavily on the TAs in the help room for their learning:

> I'll usually go home and try [to complete my homework] on my own at night, for an hour to two hours, for the two problems, usually getting a good jump on what I think the solution's going to look like. …And then, the next day, …I'll go to the tutorial room, and kind of ask the TAs that are there if my solution generally looks correct and if it looks like I've missed anything. Most of the time, there's something that I've missed, or I haven't gotten, and then I'll go sit down at the table and solve it and try to fix my solution. (Student 11)

> What I did that finally helped me understand [the course content] was immediately after class, I would just go to the help room. Just sit there; do homework where, if I needed help, I'd be able to ask the TA's there. (Student 17)

Multiple interviewees in this cluster acknowledged a tension between the expectations and value dimensions of the EV framework with regard to their peers, TAs, and instructors. For example, Student 12 said:

> I think that my experience in this class has definitely been a good one because of the tutorial room. I think that the TAs in there are an even greater resource sometimes than the professor can be because they're more available. They're more available [than the professor] and more accurate [than your peers] then you're going to get the best of both worlds there. Maybe the [group messaging with your peers] might be more available, but not as accurate. And the professor, vice-versa. (Student 12)

Overall, the perceived high availability of the TAs in the help room coupled with a perception of receiving high quality help explains why students in this cluster utilized the help room frequently.

Cluster 4 (Users of Students Not in Dynamics)

The interviews of students in this cluster did not reveal any insights into why students in this cluster utilized students not currently in Dynamics relatively frequently, but we did not explicitly ask for this information in the interviews. We did find evidence that suggests students in this cluster reached out to their peers in other sections for help (there were two sections of Dynamics each fall semester and four sections each spring semester), so students in this cluster could have misinterpreted the survey questions. The question regarding use of students outside Dynamics begins with the phrase "other students I know", and the first question asks about "my peers in the class." Thus, it is possible that students perceived the first question to be about students in their section and the subsequent question to be about students outside their section. The following is one example of how and why students worked with their peers from other sections of Dynamics and illustrates the value of aligning content across sections:

> Sometimes if I didn't understand a general topic I'd also reach out to some of my peers who were in different sections to see, "Which examples did your professor do? Did they mention anything differently?" …[For exams], I was comfortable studying with people for my exams, and also knowing that everyone was doing similar material. Not that, "Oh we have these 3 sections, and your teacher is teaching you whatever they want." The fact that it was really organized, we were

> all on the same thing at the same time, I knew that I could study with others in different sections for exams and not be at a disadvantage. (Student 26)

For students in C_4, the alignment of the Dynamics curriculum across multiple sections increased the perceived quality of the help that peers in other sections could provide.

Cluster 5 (Non-Users of the Discussion Forum, Users of Non-Dynamics Resources)

Two factors that seemed to contribute to students in this cluster not using the discussion forum were their perceptions of the forum as having a low likelihood of providing help and low quality. In certain semesters, the usage of the discussion forum was quite low, leading some students to believe that the discussion forum *could* be a useful resource, but it was not for them. For example, one student said:

> I don't think [the discussion forum is] super helpful now. I think it has the potential to be very helpful. Someone [who] took this [course] previously mentioned [that] when they took it the [discussion forum] was super, super popular. People were posting on it all the time and the only way to do the homework was to, like, look at the blog and see what people were posting. …But every time I checked, it was someone asking a question, [and] no one would answer it. Someone would ask another question; no one would answer it. Then there'd be another question that was easier that someone would answer, but it wouldn't be very descriptive. So, I think, again, it could be very, very helpful, but I don't think it was very helpful. (Student 33)

Interviewees in this cluster also did not use the discussion forum frequently because they perceived it as less helpful than other resources, indicating the value dimension of the EV model dominated their usage decision. Private communications with friends, often virtually via a group-messaging platform, were not viewable by the instructor and were perceived as being of higher quality than the public discussion forum because these private communications were more open and specific. The following quote illustrates this sentiment:

> [When asking for help on the discussion forum], it's hard to know where the line is with cheating ….[Instead,] we have an environmental engineering Dynamics [group-messaging chat], and we'll usually ask a question like, "Hey, this is what I'm doing. Does this look like what other people are doing?" Occasionally, people will post an answer like, "I got this. Is this a close answer to what anyone else

> got?" They'll get a yes or a no. Again, hard to know where that academic integrity line is, but …[the group chat is] just not as public. (Student 32)

The students in C_5 were also more likely to find other online resources outside the Dynamics learning environment. These resources included content from an online tutoring platform—a platform that concerns the instructors because of how easy it is for students to receive expedient-oriented help. However, Student 31 described the website as just another support resource that they could use if they had exhausted their other options: "…if I'm really confused and stuck I'll go on [the tutoring website] and it'll sometimes help."

Cluster 6 (Almost Exclusively Core Users)

Students in C_6 displayed similar resource-usage behaviors as C_2, but, unlike students in C_2, who often did not use the instructor's office hours because of scheduling conflicts, the interviewees in C_6 simply preferred not to use office hours. When asked how often they received help from their instructor, these two students said:

> Never. …That's just a result of me being me and not wanting to go to office hours, even though I probably should. (Student 37)

> I personally don't interact with him that much. … I think if I needed to, he would be easily reached, and I can meet with him to talk about stuff, but I personally don't. (Student 38)

It is interesting to view the above comments through the lens of the EV model. Rather than the expectations and values for the Dynamics instructor determining whether or not the students in this cluster used the instructors' office hours, the students' prior help-seeking habits seem to have dictated their decision. This result aligns with Giblin's (2016) proposition that students also use heuristics (or "empirically derived short cuts", p. 16) based on prior experience when deciding whether or not to seek help from a resource. Furthermore, portions of Student 38's quote, "I think if I needed to" and " I can meet with him", corroborate Broidy et al.'s (in review) findings that students use conditional statements and modal verbs (can/could/may/might) to justify and hedge their use of office hours.

Cluster 7 (Non-Users of Lecturebook)

We had only one interview transcript to analyze for C_7; therefore, the archetypical behaviors of this cluster as exemplified by this one student are more tentative than the other clusters. The one student we interviewed in C_7 seemed to be budget conscious, meaning the price of the lecturebook limited its accessibility. They purchased the book, but did not use it, so they returned it. Instead, the student took detailed notes when the instructor used their own slides for explaining the theory of each topic. When the instructor solved an example problem, however, they did not take notes; they listened and tried to understand the process. They took notes on how to solve the example problems later, when they watched the online solutions for the lecturebook examples in preparation for doing their homework. Because they did not have the lecturebook as a reference, they clicked through many online solution videos until they found one (or more) that looked similar to the homework problem of interest. Overall, this student did not perceive their lack of using the lecturebook as a hinderance to their learning; they had this to say about the learning process they employed: "it's less expensive, and I feel like I learn just as much."

Cluster 8 (Reliant on Peers and Lecturebook)

Students in this cluster primarily relied on their peers and lecturebook for support. All three of the interviewees exhibited similar usage behavior with the lecturebook, online videos, and their peers. They all read the theory portion of the lecturebook after the lecture to clarify the concepts, and one of the students also read it before the lecture. Overall, the interviewees found the lecturebook to be of very high value, as exemplified by this quote: "[At first] I didn't realize how helpful [the lecturebook] was and how directly it related to exactly what you're doing" (Student 35).

If working on their homework, the students in this cluster would only seek help from online example videos after they revisited the theory sections of the lecturebook. However, the students mostly used the online homework and example videos to test their understanding before an exam. One student mentioned that they consciously chose not to watch all the example videos before attempting their homework because they saved some unseen examples for their exam preparation:

> I kind of mentioned that I watch like half the lecture videos … before the homework, and then maybe I'll save them for before the exam. …Maybe I have a better chance to kind of refresh and go through new problems that I haven't seen before, right before the exam. (Student 35)

Like students in C_5, the students in this cluster appear to rely heavily on their peers for support through private channels rather than via the discussion forum. Two of the interviewees mentioned that help from people they knew and trusted was more useful than the information on the discussion forum, suggesting the quality aspect of the EV model drove their decision of who to ask for help. Unlike C_5, whose interviewees used a group-messaging platform to communicate with a larger number of students, the interviewees of C_8 kept their peer-network small. The interviewees often physically met with a small group of friends to do their homework, as described by Student 41: "If we are in a group,…we normally each do [the problem] on our own, and then stop at checkpoints, or when people get confused, and go over and make sure everyone is caught up." All of the interviewees checked their homework answers with their peers, either in person or via text messaging. The action of checking answers may appear to be performance-goal oriented, but these students described it as a way to get immediate feedback that they could use to correct misunderstandings. For example, one student said:

> Most of my homeworks, I will either work together with my roommate, or check our answers together at the end. And that's really helpful 'cause a lot of times you can figure out that you were doing something wrong, and you may have just got your math wrong, or you may have a whole concept wrong, that you probably wouldn't have caught until the homework was already graded, and you were on to the next concept in class. (Student 34)

Overall, the high perceived quality, availability, and expectation to receive help when it was needed led the interviewees in C_8 to frequently utilize a small, intimate group of peers.

Cluster 9 (Non-Users of Peers)

Most of the students in this cluster preferred to work and learn alone, which interviewees mentioned was enabled by the high availability and accessibility of the online resources. For example, one student commented:

> I don’t [interact with my peers] very often. I just like to work by myself. Especially with this class with all the resources there are online, it was easier to do that. (Student 42)

The interviewees in this cluster had varied reasons from not reaching out to their peers for help, ranging from not feeling a need for help very often to having poor experiences when they did ask for help. The following quotes illustrate this range:

> I don’t specifically meet classmates that are in my class directly. Like I said, I usually go on the [discussion forum] to see what students have written about the homework. If at all, I do have friends who are also taking the course, so I sometimes ask them about conceptual problems which don’t have solutions. (Student 43)

> I think I’ve tried [getting help from my peers] once, twice. …I found actually last year, when I tried to ask my peers for help, that everyone’s like, “if you don’t understand this by now, why are you here?” I’ve been told that a few times. I just quit [asking my peers for help]. (Student 44)

Like the quotes from C_6, the quote of Student 44 suggests that prior experiences influence current HSBs.

HSB Orientations of Each Cluster

Our qualitative data suggest that the interviewees’ HSBs were adaptive in nature across all of the resource-usage patterns and with respect to any individual resource. An example of an adaptive HSB, in the voice of the students, for each cluster is shown in Table 14. An example of how the students exhibited adaptive HSBs toward each of the nine resources considered on the survey is listed in Table 15. Evidence of all of the participants in a study exhibiting adaptive HSBs is not unprecedented; Giblin (2016) noted that all 25 of their upper-level, undergraduate math students sought help for the purpose of understanding. Because of the lack of data regarding expedient behavior, we are unable to make conclusions about the relationship between expedient HSBs and resource-usage patterns. However, the evidence of adaptive HSBs across all of the clusters suggests that all of the resource-usage patterns can support students in their desire to *understand* the content.

The final finding from the qualitative research worth mentioning is in regard to the students avoiding the instructor as a help source. Some researchers classify this type of behavior as help avoidant, (Briody et al., in review; Karabenick, 2003; Makara & Karabenick, 2013). Earlier, we posited that the relative infrequency at which students use their instructor for help could be an indication of the hierarchical power differential between instructor and student. Our qualitative results suggest that students do not use the instructor as a help source for reasons that include scheduling conflicts with office hours and personal preference, but some interviewees mentioned intimidation or the fear of being perceived as "stupid" or "dumb" as reasons for not asking the instructor questions. One quote exemplifying this sentiment was included in the C_5 section above, and another quote follows:

> Never. I mean personally I would never go to the office hours. I would rather figure it out on my own, that's just because I feel intimated almost, to go to office hours because I don't know what's going on most of the time I have to really, really struggle through problems to kind of grasp the concept. I always kind of feel stupid afterwards because I'm talking to the instructor about how I don't know anything. (Student 33)

These quotes highlight the need for instructors to actively cultivate a culture of help seeking that encourages students to ask questions and reduces their fears of "looking stupid" if they need help (Briody et al., in review; Er & Orey, 2017).

Table 14. Examples of adaptive HSBs for each cluster. The bolded text indicates key phrases related to adaptive HSBs.

Cluster	Adaptive HSB Example
C_1	We have all the homework solution videos online, so every time, if I got something wrong, **I can go online to see what steps can I improve**. (Student 4)
C_2	I remember last semester I would go to help room a lot and, like, a few semesters before that too, because there is not other, like, electronic/online resources. But for [Dynamics], I could just **take time to solve the problems by myself and understand them.** ...Usually I would just freak out by myself and just go to the help room and try to get the homework done without actually understanding [the problems], but this time I would just sit down and just try to study them. (Student 5)
C_3	I work on the problem in the Help Room, and, as soon as I get a question, I'm turning and I want an answer. Then and there while it's fresh in my mind and **I can really talk to the TA, not just about how to answer this question, but about the concept that's behind it and how I can learn from my mistake.** (Student 12)
C_4	I use my peers almost every day of the week for homework, studying, **clarification on concepts**, and they were my most useful resource. (Student 26)
C_5	Actually, several of us environmental juniors made a ... group chat. That gives us a chance to go in and say, "Hey, I'm not understanding this. **Could you please help break down this concept?**" Or "Does anyone understand what he's asking in part D of such and such problem." That gives us a chance to go back and forth and bounce ideas off each other and figure out if there's, like, competing ideas or something, and work on that. (Student 30)
C_6	I'll copy down the homework assignments, and then I'll look through the lecture example videos to see if any lecture problems that we haven't covered or similar to it, and review those. **That way I get those ideas beforehand. Then I'll try the problem.** If I can't get through it on my own, then I'll check the various blog comments that people have left, and then pitch in if I can, and then go back and just finish up the problem. The next day, after turning it in, I think, is when they put up the solutions. I guess whenever the solutions go up, I'll go back and look and see what it is that I didn't do. (Student 36)
C_7	So, for the exams, …[once] I've learned as much as I can by reading and doing my own problems, but still making mistakes, **I learn the best by explaining it to someone. Because when I explain it to someone, ...I want to make sure I'm right in how I explain it**, so then I think I subconsciously pay extra attention. That's how I learn the best, I think. (Student 40)
C_8	I guess my strategy [for completing the homework] would usually be **to watch a couple of lecture examples and go through the textbook, … then I can just go straight into the homework.** … And then as far as working with other people, I'll usually go through all of [the homework] and make sure I think I have it right and then I'll just make sure that I have the same answers hopefully, and if I don't, then I'll have to go back and redo it. (Student 35)
C_9	My preferred method [of learning] is just to go through and continually do problems. A lot of times it seems repetitious or that you're not learning anything, **but every once in a while you run into something you didn't even think of or wouldn't think to look for and then you learn that way.** (Student 42)

Table 15. Examples of adaptive HSBs for each resource listed on the survey. The bolded text indicates key phrases related to adaptive HSBs.

Resource	Adaptive HSB Example
My peers in the class	When I work with my friend, she thought of some concepts or ideas I have never thought [of] before. So, it's pretty useful to get me to understand the class materials better. And, like, **some of the homework I solve it differently, ... but in the end I try to compare my answers with her is actually proof that, like, both concepts are right.** So, they help me to understand different ways of solving the problems. (Student 5, C_2)
The course lecturebook	I do notes in all my classes the week before for the lecture, **so I'll go through what the concept that we're learning is the next week and highlight things [in the lecturebook]**, and then we go over it in class. Then I'll use it for my homework then, and then I use it to study. (Student 41, C_8)
The lecture example and homework solution videos	Yeah, I use [the homework solution videos] all the time. ... If I think I know what I'm doing **I'll go through and work the whole problem and then just skip to the end of the video to kind of see the solution and make sure I did the right thing.** (Student 31, C_5)
The course blog	…[when more students were using it], I thought [the discussion forum] was useful just **because different people brought up different things about the problems that I wouldn't think about**. (Student 6, C_2)
The instructor, by asking questions in class	I know my professor does a really good job of making sure that people actually ask questions and when they ask questions he doesn't say things like, ... "oh well that's easy, you should understand this." **He'll actually answer them and understand that not everyone understands this right away.** (Student 44, C_9)
The instructor, during office hours	My instructor, I utilized ... several times throughout the semester for office hours ... **for special clarifications on concepts** …. "That didn't click the first time. **Can you explain it to me a different way?**" Or, I'd go to a different professor's office hours to see if they explained it in a way that was better for me. (Student 26, C_4)
Online resources not accessed from the course blog (ex: online lectures or videos not associated with the course)	I do Google a whole lot of stuff. [An online tutoring website] is pretty helpful, and ... a lot of times you'll find the exact problem on [the website], but they're not that useful because the guys on [the website] get it wrong all the time. **But just to see their thought process is quite helpful.** (Student 40, C_7)
Other students I know who are not currently enrolled in the class	I do study alone, but sometimes I have friends who already took [Dynamics] before this semester, so I approach them and ask them questions ...If they still have **resources from their semesters so that I can use it and study** or just ask them for help or something. (Student 2, C_1)
The TAs in the mechanics tutorial room	So, I utilize the help room quite a bit, which there's a TA in there, but there's also other students in Dynamics who are working on the same problems. ...**I usually talk to the TA first about problems, and then once I have a good understanding of it**, if [there are] any students in that room, and there's a line for the TA**, I'll try to explain that to them** because I find I learn better when I'm explaining things. (Student 14, C_3)

Discussion

The aim of this study was to identify (RQ1) and understand (RQ2) the holistic patterns of resource usage by students in a resource-rich, blended learning environment for an undergraduate dynamics course. The summaries and implications of our results are organized below by research question.

RQ1: Patterns of Resource Usage

The most important result from the cluster analysis of the students' self-reported resource-usage data (see Table 13) was that there is not one typical resource-usage pattern for students in Dynamics; our analysis identified nine common resource-usage patterns. So, when instructors evaluate how well the Dynamics curriculum and Freeform environment supports the learning needs of all students, they should consider at least nine archetypical students, not one stereotypical student.

The identification of nine archetypical resource-usage patterns illustrates that students are tailoring their use of resources to their preferences, needs, and schedules—yet, a finite number of patterns captures how most students use the resources. The finding that all of the students are referencing multiple help sources of diverse types (face-to-face, mediated, text, video, etc.) reflects the integrated nature of the active, blended, collaborative resources of the Dynamics learning environment. When viewed through the EV framework, the fact that each of the nine resources was frequently used by at least one group of students indicates that every resource is perceived as available and valuable to at least some students, and instructors, therefore, should continue to offer the current suite of Dynamics resources in future semesters. At the same time, instructors should consider altering the resources that are less frequently used—e.g., instructor office hours—to better support more students.

The frequent use of the tutorial room by students in C_3 (the cluster with the largest membership) and C_1 indicates that access to a TA is a valued component for many students in the suite of Dynamics resources. In addition to receiving help, some students who used the tutorial room may have honed their SRL skills and enacted adaptive HSBs because, ideally, the TAs encouraged SRL and mastery-goal orientation (Puustinen, Bernicot, Volckaert-Legrier, & Baker, 2015). Student interviews suggested that some students used the tutorial room because it provided them with a more informal and less-intimidating path to expert help than visiting an instructor's office hours. Unfortunately, tutorial rooms open to all students and for specific courses are not a common resource for most universities. This study's institution makes helping students in the tutorial room the sole responsibility of the course's TAs. Because Dynamics has multiple sections and TAs are expected to work 20 hours each week, multiple TAs are hired each semester to provide 40+ hours of tutorial-room availability each week. Each instructor also hires one undergraduate student to help them grade the homework assignments. This sharing of resources across sections and the specialization of job functions allows this study's institution to offer the tutorial room, and we hope our findings encourage more engineering departments to consider providing this type of resource.

Lastly, the students' high perceptions for the expectations *and* value of the core resources was evident in the high usage frequencies for those resources. Other resources that were perceived high on only one of the dimensions of the EV framework were not used as universally. For example, non-course online resources were always available (although one could argue not as accessible because a student has to search for relevant content), but some interviewees suggested that they could be lacking in quality and accuracy. Similarly, students perceived instructors and TAs as sources of very high-quality help, but these resources were not as available or accessible as the core resources (with a possible exception being the discussion forum in semesters with low participation). In summary, our results corroborate the EV framework of Makara and Karabenick (Makara & Karabenick, 2013) and the findings of Evenhouse et al. (in preparation) and Wirtz et al. (2018) in that the availability, convenience, and quality of a resource are important factors in determining if a student will use the resource.

RQ2: Understanding How and Why

The qualitative interviews revealed a wide variety of reasons that students engaged with the resources as they did. One of the more common themes of how students utilized the Dynamics resources was to work on an assignment alone and only seek assistance if they could not overcome the challenge by themselves. The order and frequency in which they used a resource for help varied by cluster and depended on the students' perceptions of the expectations and value for each resource, as suggested by the EV model for resource selection, with the students' perceptions being influenced by their own schedule, needs, and preferences.

Overall, the results of the qualitative analysis corroborated three of the four expectation and value factors listed in Makara and Karabenick's EV model, see Figure 7. The one factor of the EV model that our interviewees never seemed to consider when choosing a resource was how well the help provided would match the type of help (adaptive or expedient) desired. However, all our interviewees described their HSBs as adaptive, so our results do not contradict the EV model, but they do not confirm the EV model either.

Implications for Practice

The cluster analysis used to group students according to how they use a set of resources is generalizable to any resource-rich learning environment. The only data required are responses to one multiple-part survey question. Thus, any instructor could employ the quantitative portion of our research design to identify how their students use the resources available to them. This information could be used to make decisions about how an instructor or department allocates their time or money to best support the learning needs of their students. For example, if administrators only looked at the average usage of each Dynamics resource (the first row of Figure 8), they may conclude that the tutorial room is used relatively infrequently compared to other lower-cost resources (like the discussion forum or the online videos). Consequently, they may decide to discontinue offering TA support in the tutorial room, or significantly cut back on staffing hours. However, the cluster analysis reveals that the students in two of the nine clusters frequently used the tutorial room, and one of those clusters had the largest membership. Thus, analyzing the behaviors of subsets of students, rather than analyzing the average behaviors of the sample, can have a practical impact on how instructors and administrators spend their time and financial resources to support students.

Because our results suggest that all of the resources can be used in adaptive ways, instructors may want to limit the time spent advising students on the specific resources they should use. Within the scope of coaching students on resource usage, instructors could reiterate the variety of resources available and emphasize the importance of the self-evaluation phase of SRL (Step 8 in Figure 6). One concrete example of how instructors could help students develop their metacognitive awareness and self-regulation is to implement a "post-test analysis," which includes multiple reflection exercises (Barkley, 2010), but there are many alternative pedagogical ideas for improving students' SRL skills in the research literature (e.g., Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010; Bandura, 1994; Linnenbrink & Pintrich, 2003; Zimmerman, 2008) Based on Karabenick and Berger's (2013) representation of help-seeking as a SRL process, as shown in Figure 6, the better students get at critically evaluating the usefulness of the help they receive, the better they should get at matching their needs to a help source. The development of SRL and help-source-matching skills would benefit all students, but it could be especially valuable for some of the students in C_1 if lower SRL skills is the cause of the students in C_1 spending the most time, on average, on Dynamics outside of class and using the most resources of any of the clusters.

Another actionable finding from this work is for instructors to consider alternatives or modifications to the online discussion forum. The cluster analysis indicates that many students use the discussion forum, but a considerable number of interviewees described it as an unreliable source of support (in terms of accuracy and expected response). Students often preferred to rely on small, private, and personal peer networks for help. Therefore, instructors should evaluate if there is a way to preserve the valued aspects of the private, smaller-group communications while also making that information available to all of the students in the class. Alternatively, Er and Orey (2017) suggested that instructor participation on the discussion forum or adding a social-networking aspect (like following, friending, or liking) could encourage participation and reducing the fear of seeking help publicly. Regardless of what modifications are considered, the asynchronous nature of the discussion forum may still cause some students to perceive it as less useful. Students who work on their homework right after it is assigned may not find the discussion forum as useful because the posted content and student participation are more limited than they are in the hours leading up to the due date (Evenhouse et al., in preparation). Nonetheless, any improvement to the content or participation on the discussion forum will likely differentially benefit those who do not have the affordances of being part of a smaller network of peers, thereby improving Freeform's ability to support the success of *all* students.

Implications for Research

This study suggests that most students seek help from multiple resources, and how and why the students choose to use certain resources varies across the resource-usage patterns. We posit that the use of multiple resources is not unique to Dynamics and is true in most undergraduate courses, especially in engineering where students often have at least a textbook (or course notes), their peers, internet resources, and the instructor's office hours available to them (Wirtz et al.). A holistic, student-centered approach allows researchers to identify and understand the multifaceted resource-usage characteristics of smaller subgroups of students, whereas the investigation of individual resources in isolation primarily reveals the average-usage statistics for each resource without contextualizing those statistics in the broader help-seeking behaviors of specific students. Therefore, to accurately understand how and why students use the resources available to them, researchers should employ a holistic, student-centered approach instead of studying the use of individual resources in isolation.

Limitations and Future Work

One of the most significant limitations of this work is the limited sample sizes of both our quantitative and qualitative data. Because the cluster analysis suggests that nine patterns describe the resources-usage patterns of most students, the number of participants (survey and interview) that we had in each cluster became limited. Future research that conducts targeted sampling of interview participants from less-common clusters would allow for a more thorough investigation into why students in those clusters engage with the resources as they do.

A second limitation of this work is that both the quantitative and qualitative analyses relied on self-reported data. For the cluster analysis, self-report errors in the resource-usage responses on the survey could have affected the students' cluster classification. We expect, however, that errors in the survey data had a minimal effect on most of the students' cluster classification because: i) the classification uncertainty for most students was very low, meaning that the resource-usage pattern of most students only aligned with that of a single cluster; ii) the nine resource-usage patterns were qualitatively unique, so gross misrepresentations of a student's resource usage would have been necessary for a student to be misclassified; and iii) the thematic analysis of interview transcripts for students within each cluster corroborated the quantitative resource-usage patterns. Regarding the qualitative data, it is possible that the interviewees did not feel comfortable sharing details of how or why they used certain resources. Some interviewees may have misrepresented their behavior because of fear for how the interviewer may perceive them (Pekrun, 2016). We tried to minimize this impression threat by conducting the interviews in a location that was not connected with Dynamics and by utilizing an interviewer who was not associated with the instructional team for Dynamics. Nonetheless, the possibility that students misrepresented their resources-usage behaviors in the interviews could have contributed to our lack of evidence regarding whether or not students enacted expedient HSBs.

In future studies, our data collection instruments and processes could be improved and aligned. The end-of-semester survey should clarify its language regarding the “blog” and students who are not currently in Dynamics. Through our interviews with students, we have found that the majority of students perceived the “blog” to be the discussion forum, but some students considered the course blog to be the course website (as a whole). In the interviews, we should explicitly ask students about their tendencies to use online resources outside of the course and peers not currently in Dynamics. We should also consider ways in which we can better research the extent to which students in each cluster exhibit expedient HSBs. One option is to consider administering a HSB instrument (e.g., see Karabenick, 2003) to get a sense of the students’ general HSB tendencies.

Finally, future work (which is presented in Chapter 4) needs to correlate the students’ resource-usage patterns to their achievement in Dynamics. The results of this study suggest that the students in all clusters sought to develop an *understanding* of dynamics. If the students across all clusters were equally successful at developing this understanding, then we would expect the performance of the students in each cluster to be similar, and a student’s resource-usage pattern would not be a significant predictor of their achievement. An insignificant relationship between a student’s resource-usage pattern and their achievement would also further strengthen the possibility that the specific resource (or usage pattern) from which a student seeks help should not be the focus of academic coaching because students can engage with the resources in many different ways and still achieve similar academic results. Alternatively, if a student’s resource-usage pattern is a significant predictor of achievement, then the proposed future work could inform the coaching of students on what resource-usage patterns might maximize their academic achievement in Dynamics.

Conclusion

As engineering continues to adopt and develop innovative teaching methods and learning environments, researchers must investigate how students experience these innovations and how the innovations affect the students’ learning. The purpose of this research was to better understand how and why students utilized the plethora of resources that were available for an undergraduate engineering course that emphasized active, blended, and collaborative learning. We identified nine qualitatively-unique patterns of resource usage, indicating that students regularly consulted multiple resources in the highly-integrated environment. Interviews suggested that the students exhibited their respective resource-usage patterns according to three out of the four factors in Makara and Karabenick's (2013) expectancy-value model for resource selection. Our interviewees described their HSBs as adaptive regardless of their resource-usage pattern; therefore, we found no evidence to support or refute the fourth factor of the expectancy-value model which contends that students will choose a help source that provides the same type of help (adaptive or expedient) that is desired. Overall, our results reflect the value of having multiple, highly-integrated resources to support students' unique needs, preferences, and adaptive HSBs. With this increased understanding of how and why students utilize the resources, instructors no longer have to rely on anecdotal evidence, assumptions, or stereotypes and can instead evaluate curricular and resource changes with regard to how the changes may affect the students of each data-driven, archetypical resource-usage pattern.

Acknowledgements

The authors would like to acknowledge the many students, staff, and faculty that made this work possible, including Garrett Stocksdale, Craig Zywicki, Rohit Kandakatla, David Evenhouse, Charles M. Krousgrill, and Angela Goldenstein. The material detailed in the present manuscript is based upon work supported by the National Science Foundation under Grant No. DUE-1525671. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 4. RELATING RESOURCE-USAGE PATTERNS TO ACADEMIC ACHIEVEMENT

Stites, N. A., Berger, E., DeBoer, J., & Rhoads, J. F. (in review). Do resource-usage patterns predict achievement?: A study of an active, blended, and collaborative learning environment for undergraduate engineering courses.

Abstract

Pedagogical innovations continue to be developed and adopted in engineering education, but how these pedagogical innovations affect the students' experiences and learning outcomes is not fully understood. This study investigates the relationship between a student's resource-usage pattern and their achievement in a resource-rich undergraduate dynamics course taught in an active, blended, and collaborative learning environment. The study extends prior research that identified nine archetypical patterns of resource usage for students in the dynamics course. A regression analysis is used to determine the extent to which a student's resource-usage pattern predicts their course grade, scores on problem-solving exam questions, and scores on conceptual exam questions. A variety of cognitive, non-cognitive, and demographic variables are included in the regression analysis to control for factors that prior research has shown are predictors of performance and could confound the resource-usage and achievement relationship. The results indicate that *on average* a student's resource-usage pattern is not predictive of their achievement, which suggests that there are many different ways to use the resources and be successful in the course. However, when individual resource-usage patterns are considered, the regression analysis identified two patterns that were associated with statistically different achievements. Students who primarily relied on their lecturebook (a custom-written textbook) and their peers for support performed higher on all three achievement metrics than their peers, and students who rarely used their lecturebook earned lower course grades and problem-solving scores. Based on the resource-usage behaviors of these two clusters, students should consider reading the lecturebook after class, but before starting their homework, and consistently collaborating with a small, intimate support group of peers. This research illustrates the power of investigating the experiences and achievements of specific subgroups of students, not just those of the stereotypical or average student.

Keywords: cluster analysis, help-seeking behavior, conceptual understanding, procedural knowledge, engineering education

Introduction

Engineering education continues to develop and implement innovative learning environments, but how these innovations affect the student experience is not fully understood. Without the knowledge of how students experience a learning environment and the resources it incorporates, instructors must rely on assumptions, stereotypes, or fragmented data to infer the students' experiences (Turns et al., 2015). The importance of knowing a product's user has been stressed by the design field for years (Cooper et al., 2007), in part because it is a way to ensure that all users' needs are considered, not just those of the majority (Rose, Harbour, Johnston, Daley, & Abarbanell, 2006). Students can be considered to be the users of educational products (the learning environment and its resources), so it is paramount to understand how the students experience a pedagogical innovation and how the innovation impacts their learning (Dean, Lee-Post, & Hapke, 2016; Minichiello & Jouffray, 2018). In this paper, we investigate the relationship between engineering students' resource-usage patterns and their academic achievement for a course that utilizes a resource-rich, active, blended, and collaborative learning environment.

In 2010, two professors of engineering developed a new learning environment for teaching an undergraduate dynamics course (Rhoads et al., 2014). This environment, named Freeform, leveraged the known benefits of active (Freeman et al., 2014), blended (Bernard et al., 2014), and collaborative (Wiggins et al., 2017) learning. In the Freeform environment, students are encouraged to take notes and solve example problems during class in a custom-written textbook, called a lecturebook, and to collaborate with their peers in class and (optionally) out of class via an online discussion forum. A suite of online solution videos (for the lecturebook examples and all of the homework problems) also supports the students' self-directed learning. Since 2010, the Freeform learning environment has been adopted for multiple engineering courses (such as statics, mechanics of materials, and vibrations), but its most mature instantiation is for an undergraduate dynamics course, which we will henceforth refer to as Dynamics, with a capital "D." The Freeform environment, and its numerous resources, offers Dynamics students many choices from which they can seek help.

Prior research employed cluster analysis to group Dynamics students according to the frequencies at which they used nine different resources: their peers, their lecturebook, online videos, an online discussion forum, their instructor during class, their instructor during office hours, online resources outside of the Dynamics learning environment, other students not enrolled in Dynamics, and a tutorial room staffed by teaching assistants (TAs; see N. A. Stites, Edward Berger, J. DeBoer, & J. F. Rhoads, in review-a). The analysis identified nine archetypical, yet qualitatively unique, patterns of resource usage. Every resource-usage pattern included the frequent use of at least two of the nine resources. The unique usage patterns involving multiple resources informs the need to study the relationship between a student's achievement and their holistic (overall) resource-usage patterns. This holistic approach is in contrast to previous studies that have only considered the average effect of individual resources on achievement (e.g., Hao et al., 2016; Horowitz et al., 2013).

The prior research on resource usage in Dynamics also detailed evidence from student interviews of adaptive help-seeking behaviors (HSBs) across all of the resource-usage patterns (Stites, Berger, et al., in review-a). Adaptive HSBs are those enacted when a student seeks understanding and autonomy, and they are generally associated with higher academic performance (Credé & Kuncel, 2008; Er & Orey, 2017; Ryan & Shin, 2011). While the prior study found evidence of adaptive HSBs across all of the resource-usage patterns, the actual academic achievements of the students exhibiting each pattern were not compared to see if the use of one particular set of resources might enable deeper understanding and better course performance than others.

Purpose of Study

The purpose of this study is to investigate the following research question: to what extent is a students' resource-usage pattern predictive of their academic achievement? We operationalize academic achievement as three measures: a student's overall course grade, their scores on problem-solving exam questions, and their scores on conceptual exam questions. We use multiple linear regression analysis to investigate the relationship between resource usage and these three measures of achievement while controlling for multiple cognitive and non-cognitive factors that can influence a student's resource usage and/or their achievement.

We utilize three measures of academic achievement because each incorporates different types of knowledge and is determined by different assessments. A student's overall course grade is an aggregate measure of a student's performance on individual and group assessments. Exam scores, meanwhile, are purely individual measures of performance. The separation of problem-solving questions from conceptual exam questions reflects the difference in format and goals of these two types of questions. The problem-solving questions assess a combination of procedural and conceptual knowledge, where procedural knowledge is knowledge of "a series of steps, or actions, done to accomplish a goal" and conceptual knowledge is knowledge of "abstract and general principles" (Rittle-Johnson et al., 2015, p. 588). The conceptual exam questions target only conceptual knowledge. Every exam in Dynamics, including the final exam, incorporates both problem-solving questions and conceptual questions because the developers of Freeform recognized the importance of both procedural and conceptual knowledge (Bohle Carbonell et al., 2014).

This research presents an alternative, holistic way of investigating the relationship between the students' resource-usage behaviors and their achievement. Prior research has analyzed students general help-seeking tendencies (e.g., Karabenick, 2003) or their use of individual resources (e.g., Horowitz et al., 2013). However, we found that most Dynamics students do not use just one resource; they use multiple resources in a variety of combinations (Stites, Berger, et al., in review-a). We expect this behavior is generally true for the students in most courses, especially undergraduate engineering-science courses and courses with many help sources available. Therefore, we argue that in many contexts the clustering of students according to their holistic resource-usage patterns is a more representative method of studying the relationship between a student's resource usage and achievement.

In general, the knowledge of how the students' resource-usage patterns relates to their academic achievement is important because it can inform how instructors coach their students to be successful in the course (Mercer, DeMaio, Wascher, Echols, & Schenck, 2018; Turns et al., 2015). If resource-usage patterns are predictive of achievement, instructors can encourage the resource usage and HSBs that are associated with higher-performing resource-usage patterns and warn against the employment of lower-performing resource-usage patterns. Instructors can also modify the course curriculum or resources to structurally encourage resource usage and HSBs that may lead to higher academic achievement. Conversely, if the students' resource-usage patterns are not predictive of academic achievement, instructors may want to limit the coaching they do on what resources the students should use, and instead, coach students on other aspects the student experience that may influence achievement, as further detailed in the conceptual framework we use.

Conceptual Framework

General Model of Learning

The conceptual framework for this study, depicted in Figure 9, is based on the motivation and self-regulated learning (SRL) model for college students that was proposed by Pintrich and Zusho (2007). Pintrich and Zusho stated that the factors that influence a student's academic achievement can be organized into five interrelated categories: personal characteristics, classroom context, motivational processes, self-regulatory processes, and past and current outcomes. The interrelated nature of this model illustrates that cognitive, non-cognitive, and demographic factors from all five variable categories can have a direct or indirect effect on achievement. For example, a student's academic major could affect how much value they see in learning dynamics, which could affect their SRL behaviors and/or their achievement (Hilpert et al., 2012; Karabenick, 2003). Another example is that a student's self-efficacy can influence their HSBs (Herring & Walther, 2016; Stump et al., 2011), and both self-efficacy and HSBs can influence performance (Karabenick, 2003; Schneider & Preckel, 2017; Stump et al., 2011; J. D. Williams & Takaku, 2011). Recursively, a student's achievement can impact their motivation, self-efficacy, HSBs, perceptions of the course content, and the instructor's behavior (e.g., Schneider & Preckel, 2017). Therefore, the model in Figure 9 indicates that we must control for many factors that can influence and confound the relationship between a single variable—e.g., resource usage—and achievement if we want to investigate a specific dyadic relationship and address limitations such as simultaneity bias. Later, we describe a regression model that we use for this purpose.

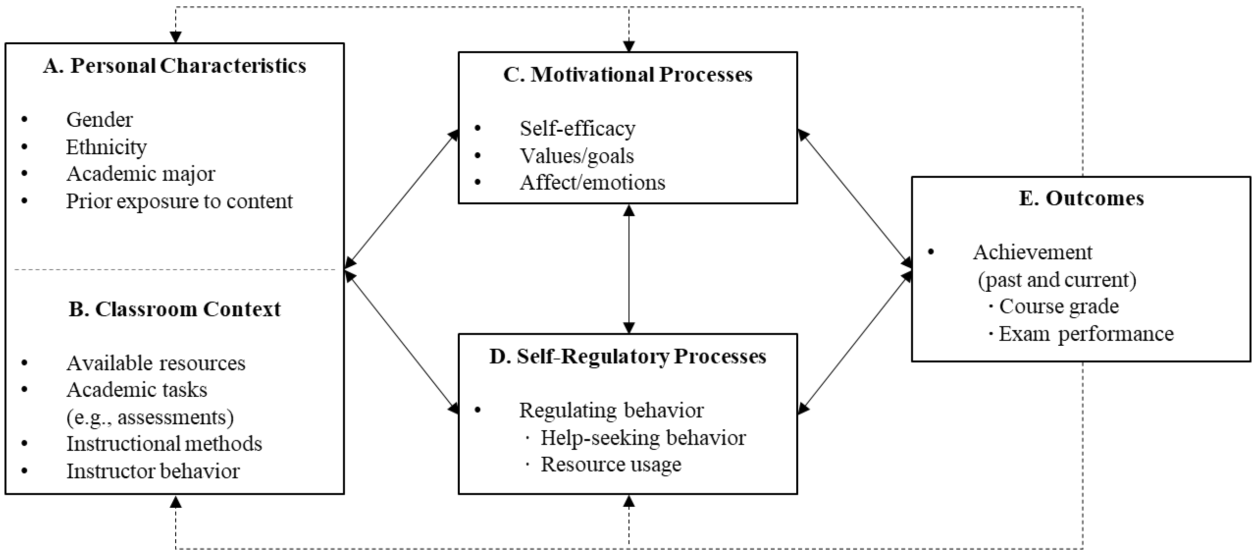

Figure 9. A model for motivation and self-regulated learning for college students. Adapted from Pintrich and Zusho (2007).

Resource Usage and Achievement

In this research, we are interested in the relationship between the students' holistic resource-usage behaviors and their achievement in Dynamics. Pintrich and Zusho did not explicitly include resource-usage behaviors in their model of learning, but we contend that resource-usage patterns are part of the "self-regulatory processes" category of variables. Resource-usage patterns are an outcome of the help-seeking process, and the help-seeking process is considered a SRL strategy (Karabenick & Berger, 2013).

Karabenick and Berger's (2013) model of the help-seeking process as a SRL strategy and Makara and Karabenick's (2013) expectancy-value model for resource selection contend that the type of help a student desires affects what resource they ask for help. The two types of help most often discussed in the literature are adaptive and expedient. Adaptive (also called instrumental or strategic) HSBs are exhibited by students who seek understanding and who have mastery-oriented goals (Horowitz et al., 2013; Karabenick, 2003; Karabenick & Berger, 2013; Newman, 2002, 2006). Expedient (also called non-adaptive or executive) HSBs, in contrast, are used by students who seek unneeded help or help that allows them to complete the task correctly with as little effort as possible (Er & Orey, 2017; Karabenick, 2003; Newman, 2006). The expedient-oriented students are usually more performance-goal oriented, meaning deep understanding of the content is a secondary concern to outperforming others. However, researchers have found that adaptive HSBs, not expedient HSBs, are generally associated with higher performance (Er & Orey, 2017; Horowitz et al., 2013; Karabenick, 2003; Ryan & Shin, 2011). In summary, because resource usage is dependent on a student's help-seeking and goal orientations, it is theoretically possible for some resource-usage patterns to be associated with students who primarily exhibit adaptive or expedient HSBs and who, therefore, have higher or lower performance, respectively.

Alternatively, Makara and Karabenick's (2013) expectancy-value model for resource selection also posits that the type of help desired is only one of the factors that can influence the likelihood of a student asking a given resource for help. Other factors that affect a student's resource selection include the student's perceptions of the availability of the resource and the quality of the help that would be provided. Therefore, it is also theoretically possible that students can have the same goal and HSB orientations but choose to use different resources because of other expectancy-value factors.

Synthesizing these theoretical scenarios and applying them to the context of this study, we would not expect any difference in achievement across resource-usage patterns if the following assumptions were true: 1) the proportion of students in each resource-usage cluster with adaptive- or expedient-oriented HSB was consistent across all of the clusters, 2) other cognitive, non-cognitive, and demographic factors that can affect achievement or resource usage were controlled for, and 3) all of the resource-usage patterns and behaviors were equally as effective in developing the students' knowledge of dynamics. Based on prior research results (see Stites, Berger, et al., in review-a) that only found evidence of adaptive HSBs across all of the resource-usage patterns, we accept the first assumption to be true. This study uses multiple regression analysis to satisfy the second assumption. The third assumption is the motivation for this study. This study aims to compare the academic achievements of similar students who choose to use different resources to determine if certain resource-usage patterns or behaviors may be more effective at developing the students' knowledge of dynamics.

Background

Dynamics Resources

Descriptions of the nine resources available to students in Dynamics are shown in Table 13. The students were asked via an online survey at the end of the semester how frequently they used each of the resources, and the median responses are also shown in Table 13. The median responses clearly indicate that the first four resources—peers, the lecturebook, online Dynamics videos, and the online discussion forum (or "blog")—are the most commonly used resources. We will refer to these four resources as the "core" resources of Freeform and further explain them below.

Much of the Dynamics curriculum and Freeform environment centers around a custom-written textbook, called a lecturebook, that concisely explains the dynamics theory and then provides many unworked example problems. The instructors and students solve many of the example problems together (with the students writing directly in their lecturebooks) during the class meeting time, which encourages students to be actively engaged in the class meetings (Chi, 2009). For each example problem in the lecturebook, and for each homework problem, a corresponding solution video exists on the course's website. Therefore, for any example problem unsolved during the class meeting, the students can watch the online videos to see how an expert approaches and solves the problem.

Peer collaboration is also in integral part of the Freeform learning environment. In class, students are often asked to complete group quizzes, with some instructors requiring that the students work with a new group for every quiz. Outside of the lecture time, an online discussion forum enables students to ask and answer questions regarding the course material and their homework problems. The Freeform philosophy encourages instructors to monitor this discussion forum and further clarify points of confusion in class but, for the most part, instructors let the discussion forum be student driven and student supported.

Table 16. A description of the nine resources included on the end-of-semester survey and the median frequency at which students used the resource (*N* = 581; adapted from Stites, Berger, et al., in review-a).

Resource	**Description**	**Median Frequency**
My peers in the class	Group quizzes in class; virtual or in-person collaboration outside of class	1-2x/wk
The course lecturebook	Combination of a workbook and concise textbook; students write notes and solve problems directly in book	3-6x/wk
The lecture example and homework solution videos	Screencasts of the instructor solving a problem; every lecturebook example and homework problem has a solution video	1-2x/wk
The course blog	"Blog" most often refers to the discussion forum, but could also be interpreted as the course website	1-2x/wk
The instructor, by asking questions in class	Could include questions before, during, or after class	1-3x/semester
The instructor, during office hours	Office hours were usually 1 hour long, 2-3 days/wk	Never
Online resources not accessed from the course blog (ex: online lectures or videos not associated with the course)	Could include online videos, online example problems, or online tutoring websites	1-3x/semester
Other students I know who are not currently enrolled in the class	Friends who have taken Dynamics previously (although there is evidence in the student interviews that students may have misinterpreted this as asking about students in other sections of the course)	Never
The TAs in the Mechanics Tutorial Room	A dedicated help room staffed over 40 hours/wk with undergraduate and graduate-student TAs	1-3x/semester

Note. The frequency-of-use data in this table represents the results from one multiple-part question from an online survey that asked students to: "Please identity how frequently you use each of the following resources for help in dynamics." The survey responses were limited to a seven-option, ordinal scale that ranged from "never" to "at least 1x/day."

Another important aspect of the Freeform learning environment is its emphasis on conceptual understanding. Instructors want the students to be capable of solving complex, mathematically-involved dynamics problems, but they also want the students to have a strong conceptual understanding of the material. Thus, every chapter of the lecturebook includes conceptual problems that the students can use to test their understanding. Videos that illustrate many of the foundational concepts of dynamics via live demonstrations are available on the course's website and are periodically shown in class. The group quizzes during class often consist entirely of conceptual questions, and conceptual problems constitute approximately one third of every exam, including the final exam. Therefore, the variety of resources that students can utilize to build conceptual understanding and procedural knowledge purposefully align with the assessments used in Dynamics.

Resource-Usage Patterns

The median frequencies for each resource delineated in Table 13 give a general sense for how frequently each resource is used when considering all of the students together. However, our prior work identified nine qualitatively-unique resource-usage patterns for Dynamics students when considering the survey responses of each student holistically in a cluster analysis. The average frequencies of use for each resource for all nine of the archetypical resource-usage patterns are shown in Figure 10. The usage statistics in Figure 10 illustrate that the students in each cluster used multiple resources for support. The students in each cluster frequently used at least two of the four core resources, so the use of all of the four core resources was not universal, as a superficial glance at Table 13 might suggest.

Our prior work also included the thematic analysis of interview data to better understand how, and why, students used the resources as they did. All of the interviewees described the type of help they desired as adaptive. Thus, all of the interviewees perceived the resource-usage pattern they employed as supporting their goal of *understanding* dynamics. We did not, however, actually compare the students' academic achievements across clusters. This study aims to determine the extent to which certain resource-usage patterns are associated with higher or lower achievement even if the students in each cluster are assumed to exhibit similar adaptive HSBs.

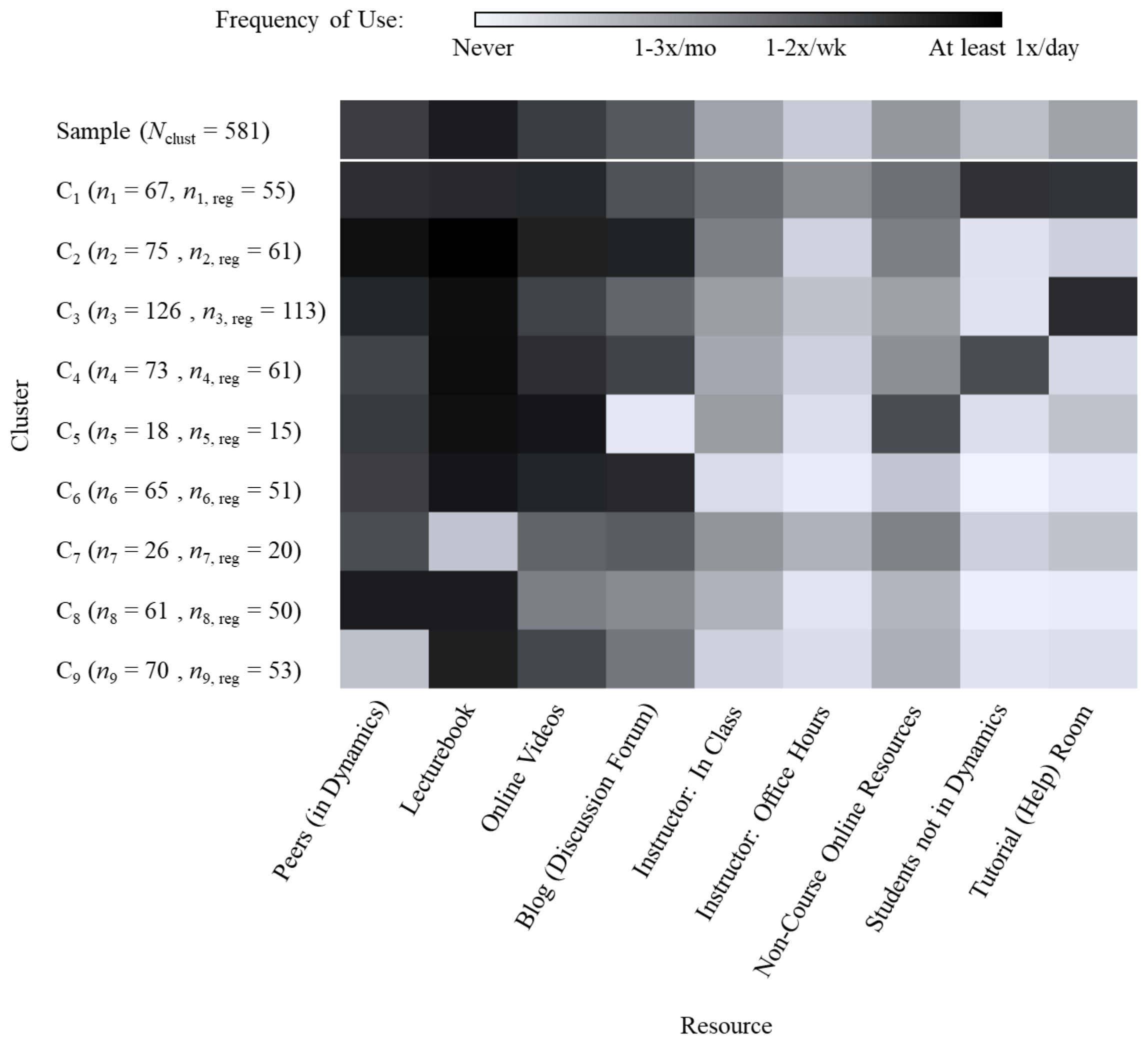

Figure 10. The Dynamics students exhibited nine archetypical resource-usage patterns (Stites, Berger, et al., in review-a). The "reg" subscript indicates the number of students in each cluster for the regression analysis, and the "clust" subscript designates the sample size for the cluster analysis. The two samples differ because of data availability.

Methods

Participants

The participants of this study were students enrolled in Dynamics at a large, public, Midwestern USA university with the highest category of research activity (Indiana University Center for Postsecondary Research, n.d.). The sampling frame for the study was all of the Dynamics students from Spring 2016 to Spring 2018 for the fall and spring semesters only. Of the 1,379 students in the sampling frame, 581 completed the survey that was used to collect data for identifying the students' archetypical patterns of resource usage. Of those students, 479 finished another survey regarding their motivations and self-efficacy and had complete registrar data, which we used for the students' cumulative GPA and demographic information. The survey on motivations and self-efficacy was not administered during the Spring 2016 semester, so no students from that semester are included in the regression analysis. There was no incentive or reward given to those students who completed the resource-usage survey during the Spring 2016-Spring 2017 semesters, but 10 points of extra credit toward the students' homework grade (which equated to less than 0.45% toward their final course grade) were offered for the Fall 2017 and Spring 2018 semesters. No incentives were offered for completing the motivations and self-efficacy survey. In summary, 581 students served as the sample for the cluster analysis that grouped the students according to their resource usage behaviors, and 479 of those students constituted the sample for the regression analysis that investigated the relationship between resource-usage patterns and academic achievement.

The demographic characteristics of these two samples are listed in Table 17, and the categories listed reflect how the institution collects and reports this data. We acknowledge that the binary representation of gender is a simplification of the gender spectrum (and that the "male" and "female" designations are classifications for sex, not gender) and that race, ethnicity, and international status are conflated together in one variable.

Table 17. The demographic characteristics of the samples used in this study. The sample used for the regression analysis is a subset of the clustering-analysis sample because of data availability.

	Clustering Sample		Regression Sample	
Variable	Count	%	Count	%
Major				
Mechanical Engineering	459	79%	376	78%
Agricultural Engineering	32	6%	26	5%
Nuclear Engineering	28	5%	20	4%
Multidisciplinary Engineering	25	4%	23	5%
Other	37	6%	34	7%
Race/Ethnicity/ International Status				
Domestic, White	379	65%	310	65%
Domestic, Asian	31	5%	27	6%
Domestic, URM	25	4%	21	4%
Domestic, Other	34	6%	31	6%
International	112	19%	90	19%
Gender				
Male	437	75%	358	75%
Female	144	25%	121	25%

Note. The sums of the percentages for the ethnicity categories for the survey participants and the major categories for the regression sample do not equal 100% because of numerical rounding.

Data Sources

In order to investigate the effect of using certain resources on a student's achievement, we had to control for as many of the other factors that can affect resource usage and achievement as possible. We used Pintrich and Zusho's model of learning, see Figure 9, as a guide for what types of control variables that we needed, and we describe the measures we used for each of the five categories of variables in the following sections.

Personal Characteristics

The data regarding the students' gender, race/ethnicity/international status, declared academic major, and whether or not they were repeating the class were gathered from the university's Registrar. Because of the small sample sizes in some of the subcategories for the race/ethnicity/international status and academic major variables, both of these variables were simplified into binary variables. The race/ethnicity/international status variable was converted into a variable that only reflected a student's international status (0 = domestic, 1 = international) to explore the differences in achievement that may be related to cultural and language differences (Kerrie A. Douglas et al., 2018; Jarvela, 2011; Ogan et al., 2014). Major was simplified to indicate whether or not a student was a mechanical engineering major (0 = mechanical engineering major, 1 = all other majors).

Classroom Context

Researchers have found that an instructor's teaching experience affects their students' course grades (Stites, Kandakatla, Berger, Rhoads, & DeBoer, in review). Students in the sections of Dynamics taught by the two developers of the Freeform environment generally earn higher course grades than students with other instructors. Stites et al. (in review) also found the number of times an instructor has taught Dynamics to be a statistically-insignificant predictor of the students' performance in the class. Therefore, for instructor-related variables, we only included a binary variable that represented if a student's instructor was a developer of Freeform or not.

All of the students in the same semester took the same exams and homework assignments, which constituted approximately 92% of a student's grade in the course. The exams and homework assignments, however, were custom written each semester, so the content and difficulty varied slightly from semester to semester. Thus, a categorical variable representing the fixed effect of each semester was used to capture this semester-to-semester variability in the assessments.

Motivational Processes

Dynamics students were asked to participate in a survey regarding, among other constructs, their perceived instrumentality (PI) of the course content and about their academic self-efficacy. We used the endogenous and exogenous PI scales (four questions for each scale) that have been included in a validation study with engineering students (Husman, Lynch, Hilpert, & Duggan, 2007) and have been used previously to study the motivations and "knowledge-building" behaviors of engineering students (Hilpert et al., 2012), which is a similar context to this study. An example of an endogenous PI question is: “What I learn in [Dynamics] will be important for my future occupational success,” and an example of an exogenous PI question is: “The grade I get in [Dynamics] will affect my future” (Hilpert et al., 2012, p. 235). The scale of academic self-efficacy consisted of eight questions and originated from Pintrich and colleagues’ (1991) Motivated Strategies for Learning Questionnaire (MSLQ). Both the PI and self-efficacy items were measured on a 1-7 Likert scale with one being “not at all true of me” and seven being “very true of me.” Negative-oriented items were reverse coded. This survey was administered in a paper format at the start of the semester (Spring 2017) or at the start and the end of the semester (Fall 2016, Fall 2017, and Spring 2018). For semesters in which this survey was administered twice, the averages of the scales’ scores were used.

Self-Regulatory Processes

The only available data we had for this category were the students’ resource-usage patterns, which were an outcome of their HSBs. We otherwise relied on variables in other categories, including the students’ cumulative GPA, to indirectly capture the students’ self-regulatory skills.

The process of clustering students according to the frequencies at which they utilized nine different resources is fully explained in Stites et al. (in review-a), and the nine archetypical resource-usage patterns are shown in Figure 10. The nine clusters (patterns) were represented with a categorical variable using weighted-effects coding so that the performance of a student within a given cluster was compared to the average performance of the students across all of the clusters, when controlling for the other variables in the model.

Outcomes (Achievement)

Because of the reciprocal nature of the learning model in Figure 9, we used the students’ cumulative GPA at the start of Dynamics (a metric of past achievement) as our final control variable and as an overall proxy measurement of factors in the other four categories of Pintrich and Zusho’s model that we did not explicitly measure. There is plenty of research to support a student’s cumulative GPA serving as a proxy measurement for these other factors. For example, Schneider and Preckel (2017) conducted a meta-analysis of prior meta-analyses on the variables associated with achievement in higher education and found that achievement is strongly related to a students' social interactions with their peers and their instructor, instructor preparation and effort, high self-efficacy, high intelligence, conscientiousness, and the use of learning strategies to achieve specific goals. Similarly, Kuh and colleagues (2006) synthesized the factor related to academic success for college students and found that a student's GPA is directly correlated to "time spent preparing for class, coming to class prepared, asking questions in class, tutoring other students, receiving prompt feedback from faculty, [and] maintaining high quality relationships with faculty" (p. 76). Overall, there is much evidence to support our use of a student's cumulative GPA as an indirect measure of personal characteristics, classroom context, motivational processes, and self-regulatory processes that may have affected the student's achievement in Dynamics.

Our outcomes of interest for Dynamics were three measures of achievement: a student's overall course grade (from the instructors' grade book, with 100% being the maximum grade), a student's total score (percent correct) on all the problem-solving exam questions, and a student's total score (percent correct) on all the conceptual exam questions. A student's overall course grade represented a weighted sum of their performance on individual and group assessments. The assessments that influenced a student's grade the most were exams, homework problems, and group quizzes, which constituted 75%, 17%, and 5% of a student's grade, respectively.

Every exam, including the final exam, included a combination of problem-solving questions and conceptual questions. The conceptual questions made up about one third of the total number of points available on the exams and were short-answer or multiple-choice questions that required little to no calculations. These questions assessed how well students understood fundamental dynamics topics conceptually, rather than how well students could solve a mathematical problem. The problem-solving questions, however, required multiple, mathematically-intense steps to solve and measured a combination of the students' procedural knowledge and conceptual knowledge (Rittle-Johnson et al., 2015). These problem-solving questions assessed the students' conceptual understanding of dynamics to an extent because to solve these multiple-step problems, the students must first choose the correct approach and the correct equation(s) of motion. Cornwell (2000) stated that students can struggle to identify the correct equation(s) of motion when they do not understand the underlying concept of the problem.

The final exam of Dynamics for all of the semesters in this study included a version of the Abbreviated Dynamics Concept Inventory (aDCI). The aDCI is a 12-item, multiple-choice instrument that is based on the 29-item Dynamics Concept Inventory (Gray et al., 2005; Stites et al., in press; Stites et al., 2016). A second version of the aDCI, aDCI.v2, was developed for Dynamics in an attempt to eliminate the construct-validity and gender-bias concerns of three aDCI.v1 items (Stites et al., in press), and was used in the semesters since Fall 2017. To account for the validity and bias issues of the aDCI.v1 items, we excluded these three items from the aDCI.v1 total scores. Thus, the aDCI.v1 scores were based on nine items, and the aDCI.v2 scores were based on 12 items. There has not been enough data collected from the students using the aDCI.v2 to conduct a second validation and fairness study, and there are not enough students in a given semester to evaluate the validity and fairness of semester-specific exam questions (problem-solving or conceptual). Overall, when possible, we evaluated the exam questions for validity, reliability, and fairness, and we excluded all questions of unsatisfactory quality from this analysis.

Data Analysis

The variables introduced above were used in a multiple linear regression analysis for each measure of achievement. However, prior to conducting the regression analysis, we explored the descriptive statistics and correlations between the variables in the model. An alpha value of 0.05 was used for determining the statistical significance of the correlation coefficients. We also compared each cluster's characteristics to those of the rest of the sample, using Fisher's exact test for count variables and Mann-Whitney U tests for the ordinal and continuous variables. To account for the threat of inflated family-wise Type I error stemming from the nine hypothesis tests (one for each cluster) for each variable, we applied a Bonferroni adjustment to each test's *p*-value and compared the adjusted *p*-value to an alpha of 0.05.

After exploring the bivariate relationships of all of the variables in the model, six multiple regression models were evaluated, two for each achievement variable. The first regression model for each achievement variable estimated how well the controls explained the variance in the achievement variable. The second model for each achievement measure included the controls and the set of variables representing the resource-usage clusters. The statistical significance, using an alpha value of 0.05, of the coefficient for each resource-usage pattern determined if the students in that cluster performed differently than the average student across all clusters when controlling for the other variables in the model. The statistical significance of the change in the adjusted R^2 value between the first and the second regression models indicated whether or not the set of resource-usage variables, as a whole, significantly improved the model fit.

Results

Descriptive and Correlation Statistics

The descriptive statistics for the control and achievement variables for the entire sample and for each cluster are shown in Tables 18-20. The correlations between the variables for the entire sample are shown in Table 21. All of the personal characteristics and classroom context variables were negatively correlated or did not have a significant relationship with each of the three achievement metrics. Conversely, the statistically-significant correlations between the achievement variables and the motivation and prior-performance variables were positive. Self-efficacy and prior GPA had the strongest correlations with achievement, which corroborates research on the predictive power of these variables (Huang & Fang, 2013; Kuh et al., 2006; Schneider & Preckel, 2017). Overall, the substantial number of statistically significant correlations in Table 21 illustrates the complex, interrelated nature of the factors that influence learning, just as Pintrich and Zusho's model in Figure 9 proposed.

Group Comparisons

When comparing the achievement of students by cluster, the results of the Mann-Whitney U tests in Table 18 suggest that the students in C_5 earned lower course grades and lower scores on problem-solving exam questions than the other students in the sample. Conversely, the students in C_8 performed better than the rest of the students on all three measures of achievement. These cluster differences in achievement are visibly evident in the distributions of the students' achievement scores for each cluster, see Figure 11. The light grey distribution lines in Figure 11 show that the achievement of the students in most of the clusters were very similar to each other and not statistically different than the sample averages.

Because the controls and achievement variables are so interrelated, it is important to view the achievement differences of Table 18 and Figure 11 in the context of any differences in the control variables. Table 19 shows that students in C_8 typically have higher self-efficacy than the other Dynamics students, students in C_2 have higher endogenous PI, and students in C_5 have lower cumulative GPAs. The correlation table, Table 21, suggested that the relationship between endogenous PI and achievement was weaker than the relationship between self-efficacy or cumulative GPA and achievement. Thus, we see no evidence of students in C_2 performing higher than their peers in Table 18, but the lower self-efficacy (although not statistically significant) and GPAs for C_5, see Figure 12, are reflected in the achievement scores of students in C_5. Similarly, the higher self-efficacy of the students in C_8 is evident in their higher performance. These findings suggest that self-efficacy and a student's prior academic performance are highly predictive of their performance in Dynamics.

Table 18. Descriptive statistics and group comparisons for the achievement metrics. Clusters with means that statistically differed (p < 0.050) from that of the rest of the sample are bolded. All p-values were multiplied by nine as a Bonferroni adjustment.

		Final Grade (%)			Problem-Solving Questions (%)			Conceptual Exam Questions (%)		
	#	Mean	SD	p-value	Mean	SD	p-value	Mean	SD	p-value
Sample	479	75.83	10.14	-	72.55	12.96	-	67.77	12.35	-
C_1	55	74.88	9.02	> 0.999	70.30	11.86	> 0.999	65.72	10.54	> 0.999
C_2	61	76.78	8.57	> 0.999	73.46	12.11	> 0.999	67.66	11.54	> 0.999
C_3	113	76.99	7.95	> 0.999	72.85	11.59	> 0.999	66.75	10.86	> 0.999
C_4	61	73.51	10.56	0.490	70.49	12.37	0.998	66.00	13.86	> 0.999
C_5	15	**65.90**	**8.90**	**0.001**	**61.31**	**12.18**	**0.005**	59.93	10.66	0.101
C_6	51	76.84	8.25	> 0.999	72.39	12.62	> 0.999	68.24	10.49	> 0.999
C_7	20	71.80	13.12	> 0.999	68.74	15.23	> 0.999	69.67	13.98	> 0.999
C_8	50	**80.63**	**10.81**	**0.002**	**78.62**	**13.48**	**0.002**	**74.50**	**13.98**	**0.001**
C_9	53	74.77	13.43	> 0.999	74.57	14.82	0.718	68.92	13.85	> 0.999

Table 19. Descriptive statistics and group comparisons for ordinal and continuous control variables. Means that statistically differed (p < 0.050) from that of the rest of the sample are bolded. All p-values were multiplied by nine as a Bonferroni adjustment.

		Self-Efficacy			PI Endogenous			PI Exogenous			Cumulative GPA		
	#	Mean	SD	p-value	Mean	SD	p-value	Mean	SD	p-value	Mean	SD	p-values
Sample	479	5.37	1.01	-	5.69	1.03	-	5.30	0.88	-	3.42	0.40	-
C_1	55	5.32	1.07	> 0.999	5.55	1.09	> 0.999	5.17	0.90	> 0.999	3.38	0.37	> 0.999
C_2	61	5.44	0.84	> 0.999	**6.03**	**0.82**	**0.037**	5.39	0.92	> 0.999	3.47	0.41	> 0.999
C_3	113	5.25	1.07	> 0.999	5.65	0.92	> 0.999	5.30	0.88	> 0.999	3.46	0.36	> 0.999
C_4	61	5.34	0.94	> 0.999	5.84	1.08	0.808	5.38	0.89	> 0.999	3.30	0.42	0.155
C_5	15	4.51	1.32	0.088	5.11	1.15	0.284	5.29	0.82	> 0.999	**3.04**	**0.26**	**0.001**
C_6	51	5.37	0.82	> 0.999	5.69	1.09	> 0.999	5.27	0.87	> 0.999	3.47	0.27	> 0.999
C_7	20	5.76	1.00	0.433	5.27	1.20	0.603	4.99	0.75	0.578	3.33	0.51	> 0.999
C_8	50	**5.76**	**0.96**	**0.006**	5.63	1.20	> 0.999	5.17	0.92	> 0.999	3.52	0.39	0.363
C_9	53	5.33	1.05	> 0.999	5.71	0.86	> 0.999	5.50	0.84	0.257	3.42	0.45	> 0.999

Note. PI = Perceived Instrumentality

Table 20. Descriptive statistics and group comparison for the dichotomous control variables. Cluster proportions that statistically differed ($p < 0.050$) from that of the rest of the sample are bolded. The p-values were multiplied by nine as a Bonferroni adjustment.

		Women		Int'l Students		Non-ME Students		Non-Dev. Instructors	
	#	Proportion	p-value	Proportion	p-value	Proportion	p-values	Proportion	p-values
Sample	479	0.25	-	0.18	-	0.22	-	0.65	-
C_1	55	0.15	0.614	**0.38**	**0.003**	0.18	> 0.999	0.60	> 0.999
C_2	61	0.28	> 0.999	0.21	> 0.999	0.16	> 0.999	0.53	> 0.999
C_3	113	0.34	> 0.999	0.14	> 0.999	0.19	> 0.999	0.49	> 0.999
C_4	61	0.26	> 0.999	0.20	> 0.999	0.18	> 0.999	0.49	> 0.999
C_5	15	0.27	> 0.999	0.13	> 0.999	0.47	0.220	0.93	0.220
C_6	51	0.27	> 0.999	0.14	> 0.999	**0.03**	**0.004**	0.44	> 0.999
C_7	20	0.10	> 0.999	0.40	0.159	0.16	> 0.999	0.64	> 0.999
C_8	50	0.16	> 0.999	**0.02**	**0.004**	0.19	> 0.999	0.60	> 0.999
C_9	53	0.26	> 0.999	0.15	> 0.999	0.23	> 0.999	0.53	> 0.999

Note. The proportions listed for Non-Dev. Instructors indicate the proportion of students who had an instructor who did not develop the Freeform environment. Non-ME represents students who declared a major other than mechanical engineering.

Table 21. Correlation coefficients (lower triangle) and their respective p-values (upper triangle) for control and achievement variables. Statistically-significant ($p < 0.050$) relationships are bolded.

	Variable	1	2	3	4	5	6	7	8	9	10	11	12
1	Gender: Female	-	**0.005**	**< 0.001**	**0.004**	0.244	**< 0.001**	**0.047**	**0.041**	**0.013**	**0.001**	**< 0.001**	**< 0.001**
2	Int'l Status: Int'l	**-0.13**	-	**0.002**	0.177	0.136	0.883	0.600	0.109	**0.017**	0.553	0.229	0.222
3	Major: Non-ME	**0.25**	**-0.14**	-	**< 0.001**	0.438	**< 0.001**	**< 0.001**	**0.029**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
4	Repeater	**0.13**	0.06	**0.19**	-	0.732	0.060	**0.043**	**0.012**	**< 0.001**	**< 0.001**	**0.011**	**0.005**
5	Instr: Non-Dev.	-0.05	0.07	0.04	0.02	-	**0.027**	**< 0.001**	0.469	**0.004**	**0.001**	0.222	**0.001**
6	Self-Efficacy	**-0.27**	-0.01	**-0.21**	-0.09	**-0.10**	-	**< 0.001**	0.541	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
7	PI Endogenous	**-0.09**	-0.02	**-0.32**	**-0.09**	**-0.15**	**0.38**	-	**< 0.001**	**0.001**	**< 0.001**	**< 0.001**	**< 0.001**
8	PI Exogenous	**0.09**	-0.07	**-0.10**	**0.11**	0.03	0.03	**0.36**	-	**0.016**	0.483	0.358	0.562
9	Cumulative GPA	**-0.11**	**0.11**	**-0.29**	**-0.28**	**-0.13**	**0.32**	**0.16**	**-0.11**	-	**< 0.001**	**< 0.001**	**< 0.001**
10	Course Grade	**-0.16**	-0.03	**-0.28**	**-0.18**	**-0.15**	**0.46**	**0.24**	-0.03	**0.75**	-	**< 0.001**	**< 0.001**
11	PS Exam Quest.	**-0.17**	0.06	**-0.22**	**-0.12**	-0.06	**0.40**	**0.16**	-0.04	**0.64**	**0.85**	-	**< 0.001**
12	Conc. Exam Quest.	**-0.22**	-0.06	**-0.24**	**-0.13**	**-0.15**	**0.47**	**0.23**	-0.03	**0.60**	**0.83**	**0.64**	-

Note. Non-Dev = Instructor who did not develop Freeform; PI = Perceived Instrumentality; PS = Problem-Solving.

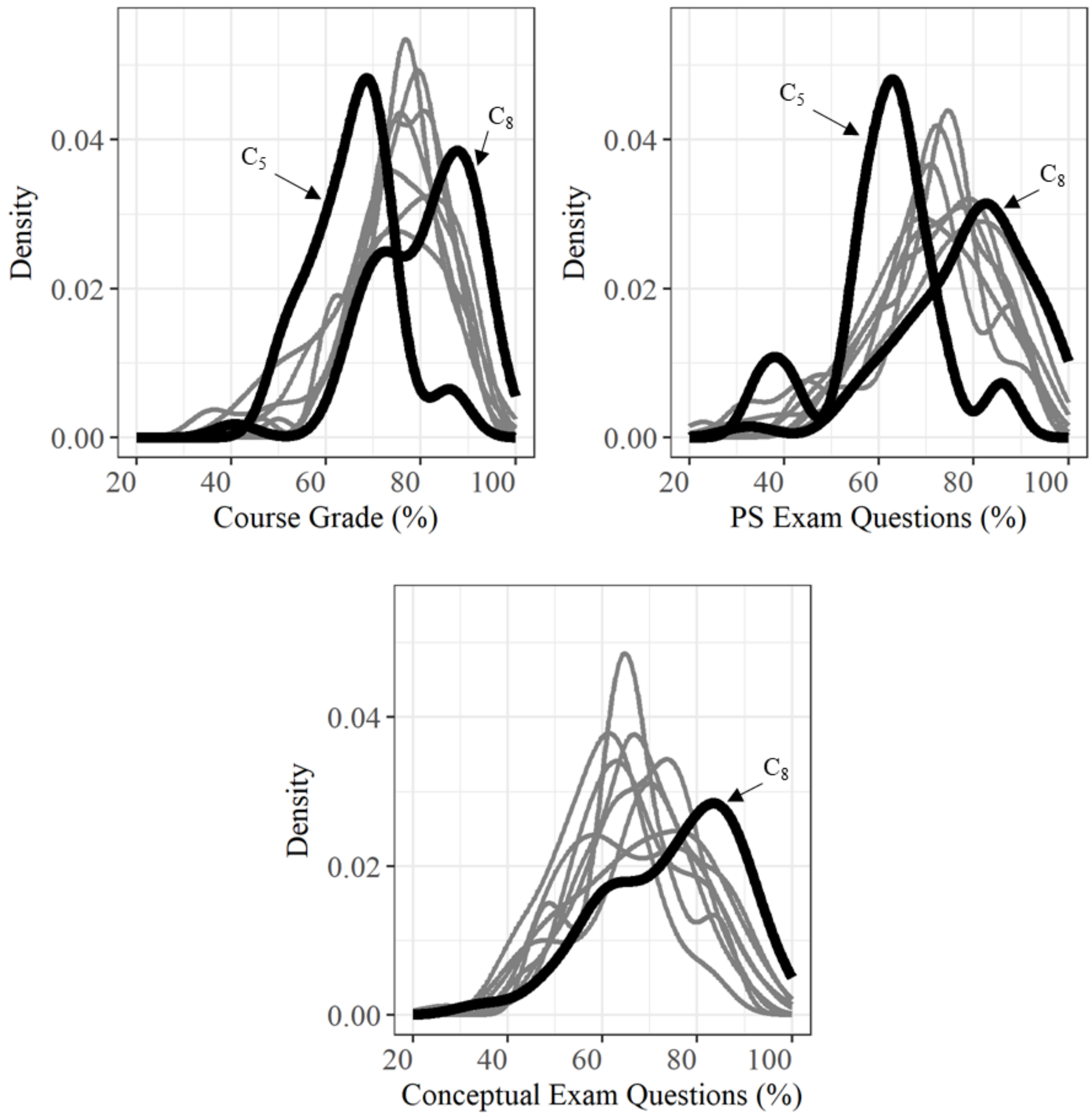

Figure 11. Cluster distributions of the three achievement metrics: overall course grade, performance on the problem-solving (PS) questions of the exams, and performance on the conceptual questions of exams. Clusters with central tendencies that differed ($p < 0.050$) from that of rest of the sample are bolded.

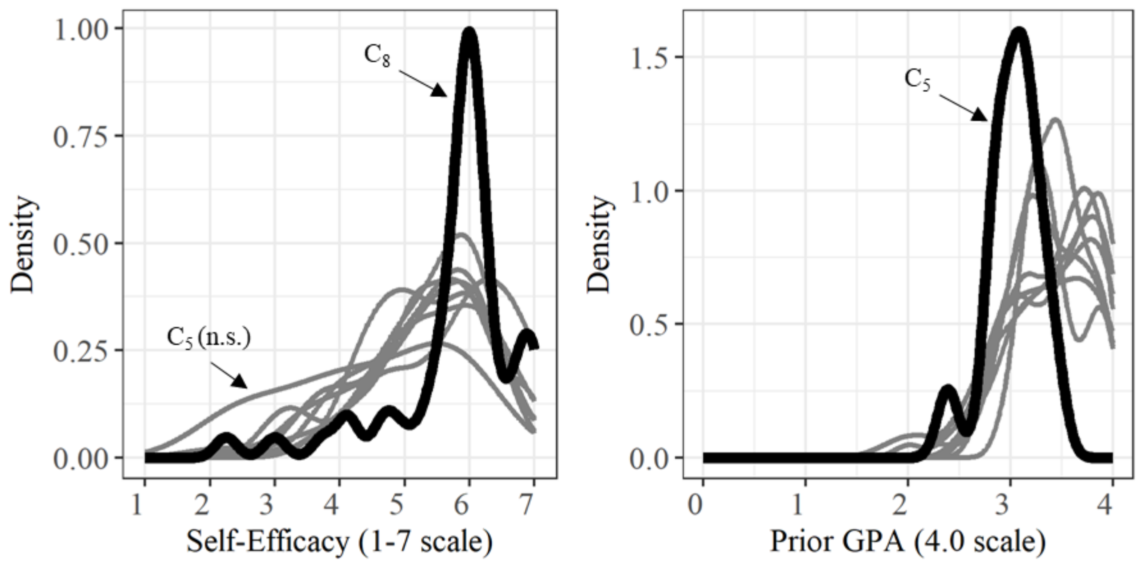

Figure 12. Cluster distributions for self-efficacy and cumulative GPA at the start of the semester in which students took Dynamics. Self-efficacy was measured on a 1-7 Likert scale that corresponded with increasing levels of self-efficacy. Cumulative GPA was measured on a 4.0 scale (A = 4.0). Clusters with central tendencies that differed ($p < 0.050$) from that of rest of the sample are bolded.

Personal characteristics and classroom context can also affect resource usage and achievement, so it is prudent to investigate the statistics of these variables by cluster as well. Table 20 shows that international students are overrepresented in C_1 and underrepresented in C_8. International students on average have higher GPAs than domestic students, so their underrepresentation in C_8—an overperforming cluster—implies that C_8 includes mostly high-performing domestic students. The overrepresentation of international students in C_1 may be indicative of international students needing to seek help from many sources because of language or cultural barriers, but the under- or overrepresentation of international students and non-mechanical-engineering majors in certain clusters warrants future study.

In summary, the results of our bivariate statistics and group-comparison tests highlight the need to control for a variety of factors that can influence a student's academic achievement, using a method such as multiple regression analysis, if we wish to isolate the relationship between resource-usage patterns and achievement in Dynamics.

Multiple Regression Analysis

The results of the multiple regression analysis are shown in Table 22. Models 1, 3, and 5 are the control models and Models 2, 4, and 6 are the full models that include the controls and the set of variables that represent the nine archetypical resource-usage patterns. Four important findings from the regression analysis are highlighted below.

First, the changes in the adjusted R^2 between the control models and the full models were not statistically significant, indicating that *on average* knowing a student's resource-usage pattern does not significantly improve the predication of their achievement in Dynamics. However, for certain students, namely those in C_7 and C_8, there is a significant relationship between their resource-usage pattern and their achievement. Students in C_7, who rarely used the lecturebook, earned lower course grades and problem-solving scores (although the latter was not quite statistically significant) than similar students in the other clusters. Students in C_8, who primarily relied on their peers and the lecturebook for help, scored statistically higher on all of the achievement metrics. We examine the ways in which the students in C_7 and C_8 used their respective resources in the Discussion section to contextualize these results.

Second, the differences in achievement for C_5 and C_8 that were evident in the group-comparison test, see Table 18, became insignificant for C_5 and remained statistically significant for C_8 after controlling for the other factors of influence. This suggests that the lower performance of C_5 can be explained by the students' personal characteristics, classroom context, motivation and self-efficacy, and/or self-regulatory processes. The higher performance of C_8, in contrast, is not fully explained by the other factors in the model. Therefore, the regression models predict that future Dynamics students who use the resources in the same way as those in C_8 will outperform similar students who exhibit a different resource-usage pattern.

Third, in a majority of the instances when a variable has a statistically-significant relationship with the scores of the problem-solving exam questions, it also has a statistically-significant relationship with the conceptual-question scores. This level of consistency likely reflects the fact that both of the outcome metrics assess a student's conceptual understanding of the material to some extent, and it also likely reflects the intertwined nature of procedural and conceptual knowledge (Rittle-Johnson & Schneider, 2015). However, at the same time, the instances in which the statistical significance of the predictor variables are not consistent across

Table 22. The estimated regression coefficient for models comparing academic achievement across clusters. Estimates that were significant ($p < 0.050$) are bolded.

	Course Grade (%)				Procedural Exam Questions (%)				Conceptual Exam Questions (%)			
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Variable	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
A. Personal Characteristics												
Gender: Female	-1.25	0.086	-1.32	0.071	**-2.44**	**0.027**	**-2.53**	**0.025**	**-4.04**	**< 0.001**	**-3.73**	**< 0.001**
Int'l Status: Int'l	**-3.33**	**< 0.001**	**-2.97**	**< 0.001**	**-2.12**	**0.075**	-1.86	0.139	**-4.39**	**< 0.001**	**-3.99**	**0.001**
Major: Non-ME	-1.13	0.165	-1.13	0.167	-0.78	0.528	-1.18	0.349	-0.68	0.554	-1.03	0.383
Repeater	3.42	0.051	**3.62**	**0.040**	4.99	0.061	5.06	0.064	**7.81**	**0.002**	**7.81**	**0.002**
B. Classroom Context												
Instr.: Non-Dev.	-1.14	0.082	-1.05	0.110	0.21	0.829	0.23	0.816	-1.34	0.153	-1.55	0.097
Academic Period												
Fall 2016	-0.19	0.863	-0.11	0.920	3.07	0.063	2.81	0.090	2.17	0.161	2.28	0.143
Spring 2017	**1.31**	**0.006**	**1.35**	**0.005**	**1.43**	**0.048**	**1.58**	**0.032**	**2.23**	**0.001**	**2.16**	**0.002**
Fall 2017	1.26	0.070	1.21	0.087	**2.58**	**0.015**	**2.67**	**0.013**	1.52	0.124	1.50	0.135
Spring 2018	**-1.21**	**< 0.001**	**-1.23**	**< 0.001**	**-2.25**	**0.000**	**-2.33**	**< 0.001**	**-2.21**	**0.000**	**-2.18**	**0.000**
C. Motivational Processes												
Self-Efficacy	**1.89**	**< 0.001**	**1.88**	**< 0.001**	**2.45**	**< 0.001**	**2.41**	**< 0.001**	**2.69**	**< 0.001**	**2.46**	**< 0.001**
PIEN	0.52	0.140	0.50	0.164	-0.33	0.541	-0.44	0.416	0.23	0.645	0.39	0.447
PIEX	0.04	0.914	0.06	0.862	0.05	0.934	-0.02	0.975	0.19	0.718	0.20	0.704
E. Prior Outcomes												
Cumulative GPA	**17.61**	**< 0.001**	**17.23**	**< 0.001**	**20.18**	**< 0.001**	**19.51**	**< 0.001**	**17.52**	**< 0.001**	**17.53**	**< 0.001**
D. Self-Regulatory Processes												
C_1			0.21	0.796			-1.82	0.141			-0.96	0.406
C_2			-0.07	0.923			0.11	0.924			-0.98	0.362
C_3			0.71	0.177			0.01	0.993			-1.21	0.104
C_4			-0.41	0.587			0.16	0.892			-0.06	0.958
C_5			-1.70	0.302			-3.17	0.206			0.81	0.730
C_6			-0.20	0.819			-0.72	0.580			-0.43	0.725
C_7			**-2.82**	**0.045**			-3.87	0.070			2.70	0.178
C_8			**1.73**	**0.046**			**2.79**	**0.034**			**3.26**	**0.008**

Table 22 continued.

C_9			-1.06	0.200			1.98	0.114			0.88	0.455
Intercept	78.10	< 0.001	77.98	< 0.001	73.83	< 0.001	73.70	< 0.001	70.91	< 0.001	70.93	< 0.001
R_{adj}^2	0.618		0.621		0.460		0.465		0.48		0.48	
ΔR_{adj}^2			0.003	0.165			0.006	0.119			0.005	0.149
				(df = 8, F = 1.47)				(df = 8, F = 1.61)				(df = 8, F = 1.52)

the achievement metrics illustrate that these three outcome variables do measure different constructs.

Lastly, the regression results highlight the complexity of the learning process. At least one variable from each category of Pintrich and Zusho's (2007) model has a statistically-significant relationship (bolded items in Table 22) with each of the three achievement measures (course grade, procedural exam questions, and conceptual exam questions). The practical implications of this result are that instructors and course designers must consider the many cognitive and non-cognitive aspects of their students when designing curricula and learning environments and researchers must control for as many of these factors as possible when investigating dyadic relationships in the learning process.

Discussion

Review of Purpose and Results

The goal of this research was to determine the extent to which a student's resource-usage pattern predicted their achievement in Dynamics. We defined achievement with three metrics: overall course grade, performance on problem-solving exam questions, and performance on conceptual exam questions. The most comprehensive tool we used to investigate this relationship was multiple regression analysis. Our regression model enabled us to control for many cognitive, non-cognitive, and demographic factors that can influence a student's learning and resource usage, as suggested by Pintrich and Zusho's model of motivation and SRL of college students. From the multiple regression analysis, we determined that the set of students' resource-usage patterns was not predictive of their achievement; however, for the students in C_7 and C_8, their resource-usage pattern was significantly related to their achievement. In the sections below, we draw on the results of our prior qualitative analysis of interview data from each cluster (see Stites, Berger, et al., in review-a) to further explore the resource-usage behaviors of C_7 and C_8. We extract suggestions on resource-usage behaviors that would be applicable to all of the students in Dynamics on how they might leverage the available resources to improve their performance in the course.

Resource-Usage Behaviors of C_8

Students who exhibited the resource-usage pattern of C_8 on average outperformed similar students who chose a different resource-usage pattern on all of the three measures of achievement. As Figure 10 illustrates, the students in C_8 primarily relied on their peers and the lecturebook for support. As discussed in our analysis of interviews from the students in each cluster (Stites, Berger, et al., in review-a), we found three common resource-usage behaviors of the students in C_8. First, if the interviewees did not understand a concept that was presented during a lecture, they would read the theory portion of the lecturebook after class but before starting their homework. They would use the lecturebook to clarify the concept and the homework problems to self-assess their understanding. Second, they primarily employed the online videos of example and homework problems as a tool for exam preparation by using these problems to self-evaluate their understanding of the material. Conversely, students from the other clusters most often used the example videos as a support resource for completing their homework. Lastly, instead of frequently using online videos for homework help, they heavily relied on a small, intimate group of friends for support. They would always check their homework answers with their small group of friends, either face-to-face or electronically, to get timely feedback on whether they understood how to solve the problem; and timely feedback has been shown to positively correlate with learning (Hattie & Timperley, 2007).

Using the resource-usage behaviors of C_8 to inform suggestions for the other students on how to improve their performance in Dynamics must be done cautiously because our thematic analysis of the interview data for C_8 was only based on three students. Nonetheless, one tentative suggestion to the students in Dynamics that is based on the HSBs of the students in C_8 would be to constantly self-evaluate their understanding of the material, and if they do not understand a concept, they should start their help-seeking process by (re)reading the theory sections of the book and then use the lecturebook examples or the homework problems to self-evaluate their understanding of the material. Also, the students should consider developing a small peer support network that they consistently utilize for help and for checking their understanding of the material. These suggestions are well rooted in other research evidence—e.g., see (Schneider & Preckel, 2017) for self-regulated learning strategies, (Berry, Cook, Hill, & Stevens, 2010; Lee, Mcneill, Douglas, Koro-Ljungberg, & Therriault, 2013) for textbook usage, and (Stump et al., 2011; Wiggins et al., 2017) for collaborative learning.

Resource-Usage Behaviors of C_7

The students in C_7 used the lecturebook much less frequently than their peers. Their infrequent use of such a principal component of the Freeform environment was surprising. The one student we interviewed from this cluster said they returned the book after buying it because they never used it. However, they also mentioned how useful it would be to have a copy of the problem statements for the example problems that the instructor solved in class, not knowing that all of the example problems used during the lecture were also in the lecturebook. Because they did not have the lecturebook, this interviewee said they took especially careful notes when the instructor was explaining the theory and conceptual foundation of a dynamics topic. Notedly, C_7 did not score significantly different than average on the conceptual exam questions. Overall, when considering the lower achievements of the students in C_7, who rarely used the lecturebook, and the higher achievements of the students in C_8, who constantly referenced the theory sections of the lecturebook, the simplest suggestion for all of the Dynamics students is to read and otherwise utilize the lecturebook.

Suggestions from Other Clusters

While the Dynamics achievements of the students in C_1 and C_5 are not statistically different than the rest of the sample when controlling for the other factors in our regression model, they still warrant discussion. First, the students in C_1 performed the same as similar students in other clusters, but they spent statistically-significantly more hours per week (M = 10.3, Med = 9, SD = 5.2) working on assignments or studying for Dynamics than many of their peers—specifically those in C_6 (M = 7.8, Med = 7, SD = 3.0), C_8 (M = 7.0, Med = 6, SD = 3.5), and C_9 (M = 7.2, Med = 6, SD = 3.4), as reported in Stites et al. (in review-a). This contrast in hours spent outside of class on Dynamics may suggest that students in C_1 could further develop their SRL skills so that they can better match their needs to the help source, or it may be indicative of beneficial characteristics like grit and persistence. Future studies should try to further understand why students in C_1 use so many resources and spend so much time on Dynamics (relative to many students) to determine how they might become more efficient with their work.

Second, the fact that students in C_5 performed the same as similar students in other clusters does not negate the fact that C_5 has a higher concentration of students who performed lower than average in Dynamics and have lower than average cumulative GPAs. According to Pintrich and Zusho's (2007) learning model, these statistics may indicate lower levels of motivation and/or self-regulation skills, which likely affect other important outcomes for these students in addition to achievement, such as persistence to graduation (Geisinger & Raman, 2013). Therefore, interventions and coaching that target those areas may be especially helpful for the students in C_5.

Implications for Practice

The results of this study suggest that on average a student's resource-usage pattern is not predictive of their performance in the class. We expect that this result is true for most courses that have high-quality and well-aligned resources. Therefore, instructors of courses with many quality resources may want to refrain from suggesting that the students use a specific resource in a specific way. For example, instructors might not want to tell students that they should use online videos for help with their homework because some students may prefer to work through a problem with their peers or consult the textbook. However, our research does suggest that there are some general resource-usage suggestions that may help students learn the course content. Students should consider mimicking the peer-collaboration strategies of C_8 by finding a small group of friends with whom they can consistently interact. In addition, students should continuously self-evaluate their understanding of the material and seek help when necessary.

The statistical significance of a student's self-efficacy and prior GPA, which we used as a proxy measurement for a student's motivational and self-regulatory processes, align with many previous studies of higher education (e.g., Robbins et al., 2004; Schneider & Preckel, 2017; Stajkovic, Bandura, Locke, Lee, & Sergent, 2018) and for engineering (e.g., French, Immekus, & Oakes, 2005; Huang & Fang, 2013; Vogt, Hocevar, & Hagedorn, 2007). These findings suggest that instructors should consider interventions that help improve a student's self-efficacy and self-regulated learning skills. Practical interventions could include: i) coaching students on the importance of time and practice so that they develop a growth mindset rather than a fixed mindset (Dweck, 2006), ii) implementing a post-test analysis that develops the students' metacognitive awareness and self-regulation (Barkley, 2010), and iii) incorporating cooperative activities and assessments into the curriculum to promote positive self-evaluations and self-efficacy (Bandura, 1994).

Implications for Research

The results of this study suggest that researchers should critically evaluate single-resource studies—i.e., studies that only consider the relationship between a student's learning and the use of a single resource, such as online videos (e.g., A. E. Williams, Aguilar-Roca, & O'Dowd, 2016) or a discussion forum (e.g., Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015). We found no evidence of a student using only a single resource; the students in all nine archetypical patterns of resource usage incorporated the frequent use of at least two resources. We posit that students utilize multiple resources (including their peers) in most of their undergraduate courses, which leads us to question the authenticity and validity of single-resource studies. Furthermore, the higher performance of the students in C_8 and the unique ways in which they used only two of the nine resources illustrates the need to understand how and why the students use, or do not use, certain resources, rather than only looking at how frequently the students use an individual resource.

We contend that our student-centered research design is generalizable to any resource-rich course or learning environment. The cluster analysis to find the students' archetypical resource-usage patterns requires survey responses from students regarding their resource-usage behaviors, and the regression analysis requires information about the control variables (collected from a survey or from the Registrar's office) and at least one outcome variable. None of these data types are unique to Dynamics.

Limitations and Future Work

One limitation of this study is the small number of control variables that were explicitly measured. Our use of a student's cumulative GPA as a proxy for the unmeasured constructs relating to personal characteristics, classroom context, motivation, and self-regulatory processes is well supported by Pintrich and Zusho's (2007) learning model and other research (e.g., Credé & Kuncel, 2008; Karabenick, 2003; Schneider & Preckel, 2017; Stajkovic et al., 2018; Wingate & Tomes, 2017). However, to better understand how other cognitive and non-cognitive factors influence achievement and resource-usage patterns, future studies could explicitly measure more control variables. Possible suggestions include classroom climate, instructor support, general HSB tendencies, and self-regulated learning skills.

Another limitation of this study was the relatively small sample sizes of some of the clusters, especially for C_5 and C_7. The small cluster sizes limited the statistical power of our regression methods. Low sample sizes also affected the qualitative analysis of our companion study which we utilized to elicit suggestions of how students should or should not use the available resources. Therefore, this study would benefit from more quantitative and qualitative data.

The generalizability of our results is limited by the fact that we only studied the resource-usage behaviors of students in one specific course, Dynamics. Nonetheless, we expect that our results are relevant to other technical courses, including engineering-science courses, with similar resources. An undergraduate dynamics course is only one of many core engineering-science courses aiming to develop the students' fundamental engineering knowledge and skills. Other courses that often have similar goals and resources (e.g., peer support, instructor's office hours, textbooks, supplemental online videos) include statics, mechanics of materials, thermodynamics, fluid mechanics, and circuits. Therefore, we posit that the generic study-skill suggestions that we developed based on the results of this study are applicable to students in many, if not most, engineering-science courses and other courses with a technical emphasis and similar resources.

Future studies should further investigate the experiences of students in minority demographic groups. The regression analysis indicated that almost all of the achievement measures were statistically-significantly lower for women and international students. The group-comparison tests indicated that international students and non-mechanical engineering students were over and underrepresented in certain resource-usage patterns. It is important to understand why these differences exist to ensure that the curriculum and learning environment provide fair and supportive experiences for all of the students, not just the majority.

Conclusion

As new learning environments and educational resources are developed for engineering education, it is important to understand how these innovations affect the students' experiences and learning. This study investigated the relationship between the students' holistic resource-usage patterns and their achievement in an undergraduate dynamics course that was taught in an active, blended, and collaborative learning environment. To our knowledge, this is the first help-seeking study to incorporate the fact that most students utilize multiple resources for help and to group student according to a cluster analysis of their self-reported usage data. Our results suggest that *on average* there is no difference in course grades, performance on problem-solving exam questions, or performance on conceptual exam questions across students who exhibited nine, qualitatively-unique, archetypical resource-usage patterns after controlling for many cognitive and non-cognitive factors that are known to influence learning. This null result suggests that in general, an instructor should limit how much time they spend on coaching students on what resources they should use. Instead, coaching and interventions that target self-efficacy or self-regulation processes may be more impactful.

However, this study also illustrates the power of considering specific users, not just the stereotypical or average users, because achievement differences were evident for students in two of the resource-usage patterns. From the resource-usage behaviors of the clusters with higher (C_8) and lower (C_7) achievement metrics, we extracted general suggestions for how students might use the resources in a manner that improves their achievement in Dynamics. These suggestions included reading the theory portion of their lecturebook, especially after a student self-determines that they do not understand a concept and before they start their homework. Also, the cultivation of a small, strongly-connected group of peers in which students consistently support the learning of everyone in the group could be beneficial. These suggestions, while stemming from this study's results, are well supported by broader research findings and, therefore, are applicable to students in other engineering-science and technical courses similar to Dynamics. Overall, data-driven knowledge of how specific subgroups of students use the course resources and perform in the class can help the instructors better coach and support the success of all students in Dynamics and negate the need to rely on assumptions or stereotypes about the students' behaviors and achievements.

Acknowledgements

The authors would like to acknowledge the many students, staff, and faculty that made this work possible, including Garrett Stocksdale, Craig Zywicki, Rohit Kandakatla, David Evenhouse, Charles M. Krousgrill, and Angela Goldenstein. The material detailed in the present manuscript is based upon work supported by the National Science Foundation under Grant No. DUE-1525671. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 5. CONCLUSION

Review of Purpose and Results

The purpose of this dissertation was to explore students' experiences and achievements in an undergraduate dynamics course that was taught in the Freeform learning environment. The two specific areas of concentration were the students' performance on an Abbreviated Dynamics Concept Inventory (aDCI) and their resource-usage behaviors. Three studies organized this work and explored: 1) the quality—validity, reliability, and fairness—of the aDCI to measure the students' conceptual understanding of Dynamics; 2) the archetypical patterns of resource usage and why the students' exhibited their respective behaviors in an active, blended, and collaborative Dynamics course; and 3) the relationship between the students' holistic resource-usage patterns and their achievements in Dynamics. The focus of each study was to employ analytical methods that would help understand the experiences and achievements of all of the students, and not just the majority.

The three studies of this dissertation utilized analytical methods that disaggregated the data according to the students' gender or holistic resource-usage pattern. The first study, as presented in Chapter 2, used multiple-group confirmatory factor analysis to test for differential item functioning (DIF) of the aDCI across genders. Chapter 3 detailed the cluster analysis of the students' holistic resource-usage behaviors to find archetypical patterns of resource usage, rather than relying on the "average" usage pattern of the entire sample. It also focused on how the students used various combinations of resources, instead of concentrating on the usage characteristics of individual resources, as previous research has done. These clusters of students according to their holistic resource-usage behaviors then served as the organizational framework for the thematic analysis of student interviews, and they were represented as a categorical variable in the regression analysis of Chapter 4 that investigated their relationship to the students' achievement in Dynamics.

The results highlight the importance of intentionally studying the experiences of smaller subsets of students when the students are delineated according to demographic or behavioral (like resource-usage) characteristics. Chapter 2 found that two of the 12 items of the aDCI items exhibited slight bias against women. Because the aDCI constituted part of the final exam for

Dynamics, this item-level bias likely had unfair practical consequences on the final-exam scores and the course grades of some women, and this bias would have gone undetected if the data were only evaluated in aggregate.

Chapter 3 determined that students exhibited nine, qualitatively-unique patterns of resource usage, none of which matched the pattern of the entire sample in aggregate. Interviews with students in each resource-usage cluster were analyzed separately, giving the experiences of each cluster equal weight and importance in the results, regardless of the number of students in each cluster. This method of analysis revealed that the students in different clusters used the resources differently because they had different perceptions about the expectations and values for certain resources.

In Chapter 4, I took a cluster-centered approach to representing resource-usage in the regression analysis. Previously, researchers had used a resource-centered approach, so their results were primarily shaped by the average use of a given resource across all students. The achievement differences found that for students in two, smaller clusters most-likely would have gone undetected with a resource-centered approach that operated on aggregated data because the results in Chapter 4 indicated that, *on average*, a student's resource-usage pattern is not predictive of their achievement in Dynamics. Also, o*n average*, the students in Dynamics frequently used their peers, the lecturebook, the online videos, and the online discussion forum. However, the students in one of the clusters that had a resource-usage pattern that was predictive of achievement rarely used the online videos or the discussion forum (an overperforming cluster), and the students in the other cluster with a significant relationship with achievement rarely used their lecturebook (an underperforming cluster). The unique relationships between resource usage and achievement for these two clusters of students would have been overshadowed by the average relationships between the variables if the analysis would have utilized aggregated data.

The results of Chapter 4 also highlight the need for more validation and fairness studies of engineering curricula, like that in Chapter 2. Demographic characteristics (gender and international status) were statistically significant predictors of achievement, even after controlling for multiple cognitive and non-cognitive factors that can influence learning. However, because little is known about the fairness of the Dynamics curriculum, it is unknown if these differences in performance are a reflection of true differences in knowledge or of bias in the assessments. Furthermore, international students were underrepresented in the one cluster that had statistically-higher performance than the other students in the course, and they were overrepresented in the cluster that spent the most hours per week working on Dynamics outside of class. These achievement and behavioral differences may have short- and long-term effects on the academic and professional lives of women and international students and, therefore, should be further investigated.

Contribution

This dissertation contributes to the scholarly advancement of knowledge in two areas: the assessment of conceptual understanding and the study of academic help-seeking behaviors. In Chapter 2, I presented a framework for validating concept inventories (CIs) that utilizes Kane's argument-based approach to validation and Messick's framing of validation as a series of hypothesis tests (Kane, 1992; Messick, 1990). This framework is very adaptable to other validation studies, and in its current form brings the often-neglected issue of fairness to the forefront. Regarding resource usage, I presented an accessible way to determine students' holistic archetypical resource-usage behaviors based on self-reported usage data. I also developed an alternative way of representing resource usage—according to the students' holistic patterns—when relating it to achievement. Previous studies have only considered general help-seeking tendencies or the use of individual resources as prediction variables, rather than considering the different combinations of resource usage that students employ. These contributions are further discussed below.

Contribution 1: An-Argument-Based Validation Approach for CIs

Concept inventories are becoming increasingly popular in engineering education as more instructors embrace the importance of students understanding the material conceptually. However, the evidence regarding the quality of these CIs varies greatly. Thus, further research on the validity, reliability, and fairness of these instruments is sorely needed to ensure that only valid inferences are drawn from the students' scores.

This need for validation studies has led to the development of evaluation frameworks for CIs (e.g., Jorion et al., 2015), but these frameworks can be very prescriptive—"do 'this' and then interpret your results according to 'this'." The argument-based validation framework that I implemented in Chapter 2 enables the researcher to decide what claims they want to make about the CI scores and what evidence is necessary to support those claims. This more-flexible approach empowers the researcher to formulate their own study, rather than prescribing a study for them.

Contribution 2: Awareness and Template for Evaluating the Fairness of CIs

The validation study in Chapter 2 also emphasized the evaluation of the aDCI for fairness, but fairness has previously been absent in the validation studies of engineering CIs. Furthermore, my results highlight the importance of considering item-level differences that can reflect biases in individual items that may otherwise be overshadowed by the average fit of a psychometric model. I focused on the gender fairness of the aDCI, but the methods I used provide a template for how to conduct similar studies of fairness for other subgroups of students.

Contribution 3: Holistic, Cluster-Based Approach to Studying HSBs

The work of Chapter 3 represents a new way of analyzing students' help-seeking behaviors. Rather than studying students' help-seeking behaviors via a survey on general help-seeking tendencies and/or investigating the students' use of individual resources (in isolation), my study clusters students according to their holistic resource-usage behaviors with course-specific resources. My holistic, student-centered approach recognizes the fact that most students do not use just one help source; they use multiple resources in different combinations, and the results in Chapters 3 and 4 illustrate that there is value to grouping students according to these combinations to understand their behaviors and achievements. Also, this course-specific scoping of the research methods gives instructors extremely specific information about the help-seeking behaviors of students in their course. The instructors can use this data-driven information to better coach the students or modify the course's resources.

Implications

This work has an impact locally, on the instructors and students in Dynamics, and on the larger engineering-education community. Locally, the validation and fairness results have already informed the development of a second version of the aDCI. The two questions that were identified as being slightly biased against women were modified and included in the second version. Future research should evaluate the impact that the changes to the content and contexts of these questions had on their gender fairness. More broadly, my work illustrates the importance of researchers and instructors evaluating the fairness of their curriculum, including CIs or other assessments, because of the considerable practical consequences that unfair curricula and assessments can have on the learning experiences and grades of specific groups of students.

Locally, the results of Chapters 3 and 4 should help instructors better understand how and why students use the resources in Dynamics. More broadly, the research designs of Chapters 3 and 4 consider the students' use of course-specific resources, rather than their general HSB tendencies as previous researchers have done. We argue that this course-specific information is more useful to instructors on how to better support their students' learning in a specific course.

The research designs of Chapters 3 and 4 also represent a shift from considering the aggregated statistics of individual resources in isolation to studying the behaviors of subgroups of students according to their holistic resource-usage patterns. We contend that most students use multiple resources (including their peers) in their undergraduate courses, especially engineering courses. Therefore, the holistic, student-centered approach to understanding students' resource-usage patterns is a more appropriate research method than studying the use of individual resources in isolation.

Lastly, in a local context, the regression results of Chapter 4 can help instructors better coach their students on how to be successful in the course. Most resource-usage patterns are associated with similar achievement, after controlling for other cognitive and non-cognitive factors. Therefore, instructors may want to limit the time they spend coaching students on what resources to use. Instead, the regression results suggest that interventions regarding self-efficacy or self-regulated learning skills may have more of an impact on the students' achievements, and practical interventions that instructors could incorporate into their curriculum are included in Chapters 3 and 4.

Limitations and Future Work

The studies in this dissertation represent the beginnings of multiple research pathways. Each study can be improved to address its limitations, extended to gain further insights, or adapted to explore similar areas. Each chapter discusses the limitations and future work related to its respective study, and I highlight a few of these recommendations below.

One of the limitations across all three of the studies is the sample sizes of the students in minority groups. In Chapter 2, the unbalanced sample sizes of men and women limited the statistical power to detect performance differences between the groups. In Chapters 3 and 4, the sample sizes of some of the resource-usage clusters were relatively small, both for the quantitative and qualitative analyses. The collection of more data for future analyses would help alleviate these limitations, especially if the sampling and recruitment procedures placed an emphasis on getting increased participation from the smaller student groups.

The validation study in Chapter 2 would also benefit from a qualitative study that aims to better understand how men and women perceive, interpret, or otherwise experience any gender biases that may be in the aDCI, the Dynamics curriculum, or the Freeform learning environment. Like Smith and Gayles (2018) and Truong et al. (2015), the qualitative study could include semi-structured interviews regarding students' experiences with gender bias in various contexts—e.g., in general, in their academic program, and in Dynamics. Similarly, the results of the group comparisons and regression analyses in Chapter 4 suggest that a fairness study of the Dynamics curriculum across international statuses or, more generally, races/ethnicities is needed. While Purdue collects information about racial/ethnicity subgroups for domestic students, they group all international students into the same category. To better understand their experiences, a more-nuanced categorization of international students (e.g., one that captures language differences) may be necessary, but this information would have to be collected directly from the students. The fairness study across racial/ethnicity/nationality groups should include quantitative methods, like those presented in Chapter 2, but a qualitative study into how students of different ethnicities experience Dynamics would also be needed to make meaning of the quantitative results.

Overall, it is important to continue researching the experiences and achievements of students outside the majority group to ensure that the engineering curricula and learning environments are fair and supportive for both majority and minority student groups.

REFERENCES

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Bandura, A. (1994). Self-efficacy. In H. Friedman (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York, NY: Academic Press.

Barkley, E. F. (2010). *Student engagement techniques: A handbook for college faculty*. San Francisco, CA: Jossey-Bass.

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22*(3), 507-526. doi:10.1037/met0000077

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education, 26*(1), 87-122. doi:10.1007/s12528-013-9077-3

Berry, T., Cook, L., Hill, N., & Stevens, K. (2010). An exploratory analysis of textbook usage and study habits: Misperceptions and barriers to success. *College Teaching, 59*(1), 31-39. doi:10.1080/87567555.2010.509376

Bohle Carbonell, K., Stalmeijer, R. E., Könings, K. D., Segers, M., & van Merriënboer, J. J. G. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review, 12*, 14-29. doi:10.1016/j.edurev.2014.03.001

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Briody, E. K., Wirtz, E., Goldenstein, A., & Berger, E. J. (in review). Breaking the tyranny of office hours: Overcoming professor avoidance.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, NY: The Guilford Press.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. doi:10.1207/s15328007sem0902_5

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105. doi:10.1111/j.1756-8765.2008.01005.x

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71. doi:10.1119/1.12989

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2015). *Applied multiple regression / correlation analysis for the behavioral sciences* (Third ed. ed.). New York, NY: Routledge.

Coller, B. (2015). *A glimpse into how students solve concept problems in rigid body dynamics*. Paper presented at the ASEE Annual Conference and Exposition, Seattle, WA.

Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: The essentials of interaction design*. Indianapolis, IN: Wiley Publishing, Inc.

Cornwell, P. J. (2000). *Dynamics evolution - Chance or design*. Paper presented at the ASEE Annual Conference and Exposition, St. Louis, MO.

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*(6), 425-453. doi:10.1111/j.1745-6924.2008.00089.x

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Boston, MA: Pearson Education.

Dean, T., Lee-Post, A., & Hapke, H. (2016). Universal Design for Learning in teaching large lecture classes. *Journal of Marketing Education, 39*(1), 5-16. doi:10.1177/0273475316662104

DeBoer, J., Stites, N., Berger, E. J., Rhoads, J. F., Krousgrill, C. M., Nelson, D. B., Zywicki, C., & Evenhouse, D. A. (2016). *Work in Progress: Rigorously assessing the anecdotal evidence of increased student persistence in an active, blended, and collaborative mechanical engineering environment*. Paper presented at the ASEE Annual Conference and Exposition, New Orleans, LA. https://peer.asee.org/27032

Dietz, R. D., Pearson, R. H., Semak, M. R., Willis, C. W., Rebello, N. S., Engelhardt, P. V., & Singh, C. (2012). Gender bias in the force concept inventory? *AIP Conference Proceedings*, *1413*, 171-174. doi:10.1063/1.3680022

Ding, L., & Caballero, M. D. (2014). Uncovering the hidden meaning of cross-curriculum comparison results on the Force Concept Inventory. *Physical Review Special Topics - Physics Education Research, 10*(2), 020125. doi:10.1103/PhysRevSTPER.10.020125

Douglas, K. A., Fernandez, T., Purzer, S., Fosmire, M., & Van Epps, A. (2018). The Critical-Thinking Engineering Information Literacy Test (CELT): A validation study for fair use among diverse students. *International Journal of Engineering Education, 34*(4), 1347-1362.

Douglas, K. A., & Pellegrino, J. (2017). *How to evaluate published assessments for use in research and evaluation*. Workshop presented at the ASEE Annual Conference and Exposition, Columbus, OH.

Douglas, K. A., & Purzer, Ş. (2015). Validity: Meaning and relevancy in assessment for engineering education research. *Journal of Engineering Education, 104*(2), 108-118. doi:10.1002/jee.20070

Douglas, K. A., Rynearson, A., & Purzer, S. (2016). Reliability, validity, and fairness: A content analysis of assessment development publications in major engineering education journals. *International Journal of Engineering Education, 32*(5A), 1960-1971.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Ballantine Books.

Er, E., & Orey, M. (2017). Understanding help-seeking bahavior of students' in a flipped classroom: A structural equation modeling approach. In M. Orey & R. M. Branch (Eds.), *Educational media and technology yearbook*. Switzerland: Springer International Publishing.

Evenhouse, D., Kandakatla, R., Berger, E., Rhoads, J. F., & DeBoer, J. (in preparation). Choice and balance: Examining undergraduate engineering students' usage of resources in a blended learning environment.

Fang, N. (2012). Students' perceptions of dynamics concept pairs and correlation with their problem-solving performance. *Journal of Science Education and Technology, 21*(5), 571-580. doi:10.1007/s10956-011-9347-7

Fraley, C., & Raftery, A. E. (2007). Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software, 18*(6), 1-13. doi:10.18637/jss.v018.i06

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. doi:10.1073/pnas.1319030111

French, B. F., Immekus, J. C., & Oakes, W. C. (2005). An examination of indicators of engineering students' success and persistence. *Journal of Engineering Education, 94*(4), 419-425. doi:10.1002/j.2168-9830.2005.tb00869.x

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18. doi:10.1037/a0024338

Froyd, J. E., & Ohland, M. W. (2005). Integrated engineering curricula. *Journal of Engineering Education, 94*(1), 147-164. doi:10.1002/j.2168-9830.2005.tb00835.x

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Geisinger, B. N., & Raman, D. R. (2013). Why they leave: Understanding student attrition from engineering majors. *International Journal of Engineering Education, 29*(4), 914-925.

Giblin, J. (2016). *Factors influencing the selection of academic help sources*. (Doctoral dissertation). Old Dominion University, Norfolk, VA.

Gray, G. L., Costanzo, F., Evans, D., Cornwell, P., Self, B., & Lane, J. L. (2005). *The Dynamics Concept Inventory assessment test: A progress report and some results*. Paper presented at the ASEE Annual Conference and Exposition, Portland, OR. https://peer.asee.org/15378

Griffin, W., Cohen, S. D., Berndtson, R., Burson, K. M., Camper, K. M., Chen, Y., & Smith, M. A. (2014). Starting the conversation: An exploratory study of factors that influence student office hour use. *College Teaching, 62*(3), 94-99. doi:10.1080/87567555.2014.896777

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. doi:10.1119/1.18809

Hao, Q., Barnes, B., Branch, R. C., & Wright, E. (2016). *Predicting college students' online help-seeking behavior: The effect of learning proficiency, interest, prior knowledge, epistemological belief, and problem difficulty*. Paper presented at the AERA Annual Meeting, Philadelphia, PA.

Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Research and Clinical Center for Child Development Annual Report, 6*, 27-36.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi:10.3102/003465430298487

Henderson, C. (2002). Common concerns about the Force Concept Inventory. *The Physics Teacher, 40*(9), 542-547. doi:10.1119/1.1534822

Herring, C., & Walther, J. (2016). *Academic help-seeking as a stand-alone, metacognitive action: An empirical study of experiences and behaviors in undergraduate engineering students*. Paper presented at the ASEE Annual Conference and Exposition, New Orleans, LA. https://peer.asee.org/26490

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*(3), 141-158. doi:10.1119/1.2343497

Hilpert, J. C., Husman, J., Stump, G. S., Kim, W., Chung, W.-T., & Duggan, M. A. (2012). Examining students' future time perspective: Pathways to knowledge building. *Japanese Psychological Research, 54*(3), 229-240. doi:10.1111/j.1468-5884.2012.00525.x

Horowitz, G., Rabin, L. A., & Brodale, D. L. (2013). Improving student performance in organic chemistry: Help seeking behaviors and prior chemistry aptitude. *Journal of Scholarship of Teaching and Learning, 13*(3), 120-133.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. doi:10.1037//1082-989x.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education, 61*, 133-145. doi:10.1016/j.compedu.2012.08.015

Husman, J., Lynch, C., Hilpert, J., & Duggan, M. A. (2007). *Validating measures of future time perspective for engineering students: Steps toward improving engineering education*. Paper presented at the ASEE Annual Conference and Exposition, Honolulu, HI. https://peer.asee.org/3045

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90-98. doi:10.1016/j.jcps.2009.09.003

Indiana University Center for Postsecondary Research. (n.d.). The Carnegie classification of institutions of higher education. Retrieved from http://carnegieclassifications.iu.edu/

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651-666. doi:10.1016/j.patrec.2009.09.011

Jamieson, L. H., & Lohmann, J. R. (2012). *Innovation with impact*. Washington, DC: American Society of Engineering Education.

Jarvela, S. (2011). How does help seeking help? – New prospects in a variety of contexts. *Learning and Instruction, 21*(2011), 297-299. doi:10.1016/j.learninstruc.2010.07.006

Johnson, D. G. (2010). Sorting out the question of feminist technology. In L. L. Layne, S. L. Vostral, & K. Boyer (Eds.), *Feminist Technology* (pp. 36-54). Urbana, IL: University of Illinios Press.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative learning: Increasing college faculty instructional productivity. ASHE-ERIC Higher Education Report No. 4*. Washington, D.C.: The George Washington University, School of Education and Human Development.

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638-654. doi:10.1111/jcal.12107

Jorion, N., Gane, B. D., James, K., Schroeder, L., DiBello, L. V., & Pellegrino, J. W. (2015). An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, *104*(4), 454-496. doi:10.1002/jee.20104

Joyce, A. (2016). Framing office hours as tutoring. *College Teaching*, *65*(2), 92-93. doi:10.1080/87567555.2016.1189393

Kandakatla, R., Berger, E., Rhoads, J. F., & DeBoer, J. (in review). Student perspectives on the learning resources in an active, blended, and collaborative (ABC) learning environment in an undergraduate engineering course.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535. doi:10.1037//0033-2909.112.3.527

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology, 28*(2003), 37-58. doi:10.1016/s0361-476x(02)00012-7

Karabenick, S. A., & Berger, J.-L. (2013). Help seeking as a self-regulated learning strategy. In H. Bembenutty, T. J. Cleary, & A. Kitsantas (Eds.), *Applications of self-regulated learning across diverse disciplines*. Charlotte, NC: Information Age Publishing, Inc.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success : A review of the literature*. Retrieved from https://nces.ed.gov/npec/pdf/Kuh_Team_ExecSumm.pdf

Lee, C. S., Mcneill, N. J., Douglas, E. P., Koro-Ljungberg, M. E., & Therriault, D. J. (2013). Indispensable resource? A phenomenological study of textbook use in engineering problem solving. *Journal of Engineering Education, 102*(2), 269-288. doi:10.1002/jee.20011

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, *19*(2), 119-137. doi:10.1080/10573560308223

Litzinger, T., Lattuca, L. R., Hadgraft, R., & Newstetter, W. (2011). Engineering education and the development of expertise. *Journal of Engineering Education, 100*(1), 123-150. doi:10.1002/j.2168-9830.2011.tb00006.x

Madsen, A., McKagan, S. B., & Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics - Physics Education Research, 9*(2), 020121. doi:10.1103/PhysRevSTPER.9.020121

Makara, K. A., & Karabenick, S. A. (2013). Characterizing sources of academic help in the age of expanding educational technology: A new conceptual framework. In S. A. Karabenick & M. Puustinen (Eds.), *Advances in help-seeking research and applications: The role of emerging technologies*. Charlotte, NC: Information Age Publishing.

Martin, M. M., & Myers, S. A. (2006). Students' communication traits and their out-of-class communication with their instructors. *Communication Research Reports, 23*(4), 283-289. doi:10.1080/08824090600962599

Maydeu-Olivares, A., & García-Forero, C. (2010). Goodness-of-fit testing. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 190-196). Oxford, UK: Elsevier.

McCullough, L. (2004). Gender, context, and physics assessment. *Journal of International Women's Studies, 5*(4), 20-30.

McCullough, L. (2011). *Gender differences in student responses to physics conceptual questions based on question context*. Paper presented at the ASQ Advancing the STEM Agenda in Education, the Workplace and Society Conference, Menomonie, WI.

McKenna, A. F. (2007). An investigation of adaptive expertise and transfer of design process knowledge. *Journal of Mechanical Design, 129*(7), 730-734. doi:10.1115/1.2722316

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1-47.

Mercer, D. M., DeMaio, M. A., Wascher, D. C., Echols, P. G., & Schenck, R. C. (2018). Teaching as coaching. In P. J. Dougherty & B. L. Joyce (Eds.), *The orthopedic educator* (pp. 61-76). Basel, Switzerland: Springer International Publishing AG.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543. doi:10.1007/bf02294825

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Franscisco, CA: Jossey-Bass.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. doi:10.2307/1175249

Messick, S. (1990). *Validity of test interpretation and use*. Princeton, NJ: Educational Testing Service.

Minichiello, A., & Jouffray, M. (2018). *Communicating findings about online forum use among undergraduates in distance-delivered calculus: Developing a help seeking usage model*. Paper presented at the ASEE Annual Conference and Exposition, Salt Lake City, UT. https://peer.asee.org/29869

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6-20. doi:10.1111/j.1745-3992.2006.00075.x

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment* (J. W. Pellegrino, N. Chudowsky, & R. Glaser Eds.). Washington, DC: National Academies Press.

National Science Board. (2016). *Science and engineering indicators 2016*. Arlington, VA: National Science Foundation (NSB-2016-1).

National Science Board. (2018). *Science and engineering indicators 2018*. Alexandria, VA: National Science Foundation.

Newman, R. S. (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory Into Practice, 41*(2), 132-138. doi:10.1207/s15430421tip4102_10

Newman, R. S. (2006). Students' adaptive and nonadaptive help seeking in the classroom: Implications for the context of peer harassment. In S. A. Karabenick & R. S. Newman (Eds.), *Help seeking in academic settings: Goals, groups, and contexts*. New York, NY: Routledge.

Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2014). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education, 25*(2), 229-248. doi:10.1007/s40593-014-0034-8

Osborn Popp, S. E., Meltzer, D., & Megowan-Romanowicz, M. C. (2001). *Is the Force Concept Inventory biased? Investigating differential item functioning on a test of conceptual learning in physics*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Pandy, M. G., Petrosino, A. J., Austin, B. A., & Barr, R. E. (2004). Assessing adaptive expertise in undergraduate biomechanics. *Journal of Engineering Education, 93*(3), 211-222. doi:10.1002/j.2168-9830.2004.tb00808.x

Pekrun, R. (2016). Using self-report to assess emotions in education. In M. Zembylas & P. A. Schutz (Eds.), *Methodological advances in research on emotion and education*. Cham, Switzerland: Springer.

Physport. (n.d.). Retrieved from https://www.physport.org/assessments/FCI

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: University of Michigan Press.

Pintrich, P. R., & Zusho, A. (2007). Student motivation and self-regulated learning in the college classroom. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective*. Dordrecht, Netherlands: Springer.

Prebel, T., Stites, N. A., Berger, E., Rhoads, J. F., & DeBoer, J. (2017). *Work in Progress: Predictive analysis of conceptual understanding based on self-reported student engagement with resources in a blended engineering class*. Paper presented at the Research in Engineering Education Symposium, Bogota, Colombia.

Puustinen, M., Bernicot, J., Volckaert-Legrier, O., & Baker, M. (2015). Naturally occurring help-seeking exchanges on a homework help forum. *Computers & Education, 81*(2015), 89-101. doi:10.1016/j.compedu.2014.09.010

Reed-Rhoads, T., & Imbrie, P. K. (2008). *Concept inventories in engineering education*. Paper presented at the Board of Science Education Workshop on the Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education, Washington, DC.

Rennie, L. J., & Parker, L. H. (1993). Assessment in physics: Further exploration of the implications of item context. *The Australian Science Teachers Journal, 39*(4), 28-32.

Rhoads, J. F., Nauman, E., Holloway, B., & Krousgrill, C. (2014). *The Purdue Mechanics Freeform Classroom : A new approach to engineering mechanics education*. Paper presented at the ASEE Annual Conference and Exposition, Indianapolis, IN. https://peer.asee.org/23174

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. C. Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. Oxford, United Kingdom: Oxford University Press.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587-597. doi:10.1007/s10648-015-9302-x

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288. doi:10.1037/0033-2909.130.2.261

Rose, D. H., Harbour, W. S., Johnston, C. S., Daley, S. G., & Abarbanell, L. (2006). Universal Design for Learning in postsecondary education: Reflections on principles and their application. *Journal of Postsecondary Education and Disability*, *19*(2), 135-151.

Ryan, A. M., & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction, 21*(2011), 247-256. doi:10.1016/j.learninstruc.2010.07.003

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in Multivariate Statistical Analysis* (pp. 233-247). Boston, MA: Springer.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210-222. doi:10.1016/j.hrmr.2008.03.003

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565-600. doi:10.1037/bul0000098

Scott, T. F., Schumayer, D., & Gray, A. R. (2012). Exploratory factor analysis of a Force Concept Inventory data set. *Physical Review Special Topics - Physics Education Research, 8*(2), 020105. doi:10.1103/PhysRevSTPER.8.020105

Self, B. P., & Widmann, J. M. (2017). *Demo or hands-on? A crossover study on the most effective implementation strategy for inquiry-based learning activities*. Paper presented at the ASEE Annual Conference and Exposition, Columbus, OH. https://peer.asee.org/28101

Shryock, K. J., & Froyd, J. E. (2011). *Alignment of preparation via first-year physics mechanics and calculus courses with expectations for a sophomore statics and dynamics course*. Paper presented at the ASEE Annual Conference and Exposition, Vancouver, BC.

Smith, K. A., Johnson, D. W., & Johnson, R. T. (1981). Structuring learning goals to meet the goals of engineering education. *Engineering Education, 72*(3), 221-226.

Smith, K. N., & Gayles, J. G. (2018). "Girl power": Gendered academic and workplace experiences of college women in engineering. *Social Sciences, 7*(1), 11. doi:10.3390/socsci7010011

Socha, A., DeMars, C. E., Zilberberg, A., & Phan, H. (2015). Differential item functioning detection with the Mantel-Haenszel procedure: The effects of matching types and other factors. *International Journal of Testing, 15*(3), 193-215. doi:10.1080/15305058.2014.984066

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society, 76*(3), 483-493.

Stajkovic, A. D., Bandura, A., Locke, E. A., Lee, D., & Sergent, K. (2018). Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: A meta-analytic path-analysis. *Personality and Individual Differences, 120*, 238-245. doi:10.1016/j.paid.2017.08.014

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306. doi:10.1037/0021-9010.91.6.1292

Stites, N. A., Berger, E., DeBoer, J., & Rhoads, J. F. (in review-a). A cluster-based approach to identifying and understanding students' archetypical resource-usage patterns in an active, blended, and collaborative learning environment.

Stites, N. A., Berger, E., DeBoer, J., & Rhoads, J. F. (in review-b). Do resource-usage patterns predict achievement?: A study of an active, blended, and collaborative learning environment for undergraduate engineering courses.

Stites, N. A., Douglas, K. A., Evenhouse, D., Berger, E., DeBoer, J., & Rhoads, J. F. (in press). A validation and differential item functioning (DIF) study of an Abbreviated Dynamics Concept Inventory. *International Journal of Engineering Education*.

Stites, N. A., Evenhouse, D. A., Tafur, M., Krousgrill, C. M., Zywicki, C., Zissimopoulos, A. N., Nelson, D. B., DeBoer, J., Rhoads, J. F., & Berger, E. J. (2016). *Analyzing an abbreviated Dynamics Concept Inventory and its role as an instrument for assessing emergent learning pedagogies*. Paper presented at the ASEE Annual Conference and Exposition, New Orleans, LA. https://peer.asee.org/26242

Stites, N. A., Kandakatla, R., Berger, E., Rhoads, J. F., & DeBoer, J. (in review). Does instructor experience matter?: A study of the relationship between instructor experience and student performance in an active, blended, and collaborative undergraduate engineering course.

Streveler, R., Litzinger, T., Miller, R., & Steif, P. (2008). Learning conceptual knowledge in the engineering sciences: Overview and future research directions. *Journal of Engineering Education, 97*(3), 279-294. doi:10.1002/j.2168-9830.2008.tb00979.x

Streveler, R. A., Miller, R. L., Santiago-Roman, A. I., Nelson, M. A., Geist, M. R., & Olds, B. M. (2011). Rigorous methodology for concept inventory development: Using the 'assessment triangle' to develop and test the Thermal and Transport Science Concept Inventory (TTCI). *International Journal of Engineering Education, 27*(5), 968-984.

Stump, G. S., Hilpert, J. C., Husman, J., Chung, W.-T., & Kim, W. (2011). Collaborative learning in engineering students: Gender and achievement. *Journal of Engineering Education, 100*(3), 475-497. doi:10.1002/j.2168-9830.2011.tb00023.x

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tavakol, M., & Dennick, R. (2011). Post-examination analysis of objective tests. *Medical Teacher, 33*(6), 447-458. doi:10.3109/0142159X.2011.564682

Traxler, A., Henderson, R., Stewart, J., Stewart, G., Papak, A., & Lindell, R. (2018). Gender fairness within the Force Concept Inventory. *Physical Review Physics Education Research, 14*(1), 010103. doi:10.1103/PhysRevPhysEducRes.14.010103

Truong, K. A., Museus, S. D., & McGuire, K. M. (2015). Vicarious racism: A qualitative analysis of experiences with secondhand racism in graduate education. *International Journal of Qualitative Studies in Education, 29*(2), 224-247. doi:10.1080/09518398.2015.1023234

Turns, J., Borgford-Parnell, J., & Ferro, T. (2015). *Exploring the usefulness of personas in engineering education*. Paper presented at the Research in Engineering Education Symposium, Dublin, Ireland.

Vandenberg, R. J., & Lance, C. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. doi:10.1177/109442810031002

Viiri, J. (2003). Engineering teachers' pedagogical content knowledge. *European Journal of Engineering Education, 28*(3), 353-359.

Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A social cognitive construct validation: Determining women's and men's success in engineering programs. *The Journal of Higher Education, 78*(3), 337-364. doi:10.1353/jhe.2007.0019

Wang, J., & Bao, L. (2010). Analyzing Force Concept Inventory with item response theory. *American Journal of Physics, 78*(10), 1064-1070. doi:10.1119/1.3443565

White, B. Y. (1983). Sources of difficulty in understanding Newtonian dynamics. *Cognitive Science, 7*(1), 41-65. doi:10.1207/s15516709cog0701_2

Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences. *AERA Open, 3*(2), 1-14. doi:10.1177/2332858417708567

Williams, A. E., Aguilar-Roca, N. M., & O'Dowd, D. K. (2016). Lecture capture podcasts: Differential student use and performance in a large introductory course. *Educational Technology Research and Development*, *64*(1), 1-12. doi:10.1007/s11423-015-9406-5

Williams, J. D., & Takaku, S. (2011). Help seeking, self-efficacy, and writing performance among college students. *Journal of Writing Research*, *3*(1), 1-18. doi:10.17239/jowr-2011.03.01.1

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.

Wingate, T. G., & Tomes, J. L. (2017). Who's getting the grades and who's keeping them? A person-centered approach to academic performance and performance variability. *Learning and Individual Differences*, *56*, 175-182. doi:10.1016/j.lindif.2017.02.007

Wirtz, E., Dunford, A., Berger, E. J., Briody, E., Guruprasad, G., & Senkpeil, R. (2018). Resource usage and usefulness: Academic help-seeking behaviours of undergraduate engineering students. *Australasian Journal of Engineering Education*. doi:10.1080/22054952.2018.1525889

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-183. doi:10.3102/0002831207312909

APPENDIX A. FULL-STUDY CONSENT FORM

Purdue IRB Protocol #: 1507016258 - Expires on: 08-JUN-2018

RESEARCH PARTICIPANT CONSENT FORM

Understanding and Supporting Mechanical Engineering Undergraduate Student and Faculty Engagement with an Active, Blended and Collaborative (ABC) Learning Environment

PI Jennifer DeBoer
Engineering Education
Purdue University

What is the purpose of this study?

You are asked to participate in a study conducted by Profs. Jennifer DeBoer, Ed Berger, Jeff Rhoads, and Chuck Krousgrill of the Engineering Education and Mechanical Engineering Departments of Purdue University. You were selected as a participant in this study because you are registered in ME 274 this semester. You should read the information below before deciding whether or not to participate. The **purpose of this study** is to understand the use and possible impact of the unique format of the Freeform classroom. This class environment includes in-class learning activities, online videos and lecturebooks that are linked with online work, and collaborative in-class and online spaces (the course blog). This information may help to improve classrooms in engineering and STEM classrooms at multiple universities.

What will I do if I choose to be in this study?

All students in participating classes will complete pre- and post-course survey instruments, which include surveys of study strategies/self-efficacy (MSLQ), concept knowledge (DCI), motivation and perceptions, and questions about your attitude about the classroom structure and course content. This also includes your previous course grade data in order to understand how students with different previous experiences go through this environment. All participating classes will have a video-recorded instructor observation 5 times per semester, while a subset of class sections will have video-recorded instructor observations of every class session. All participating faculty and students will have their engagement with the blog and Youtube videos recorded by Google Analytics. A subset of students will be invited to participate in a lab session that will include a semi-structured interview and a recorded think-aloud problem solving session that will utilize eye-gaze tracking and digital voice/audio recording. This random sample of students will be asked to participate in interviews and allow researchers to observe them doing homework problems. Participation or non-participation in the study will have no effect on your grade.

How long will I be in the study?

You will be a participant in the study for the duration of ME 274. Data analysis will continue after your enrollment in ME 274 ends. We do not anticipate asking you for any further information once ME 274 has ended this semester. Survey instruments and concept tests will, for the most part, be completed as part of your regular course.

What are the possible risks or discomforts?

Minimal possible risks or discomfort. Breach of confidentiality is the only risk and the safeguards used to minimize this risk can be found in the confidentiality section. Participation in this study will have no relationship or consequences for your grade in ME 274.

Are there any potential benefits?

There are no direct benefits to the participant. Your participation may provide you the chance to share your experiences with the new type of learning environment in ME 274. Your valuable perspective may help to improve this environment for future students.

Purdue IRB Protocol #: 1507016258 - Expires on: 08-JUN-2018

Will information about me and my participation be kept confidential?

The project's research records may be reviewed by the National Science Foundation and by departments at Purdue University responsible for regulatory and research oversight. Any information obtained in connection with these data collections and that may be identified with you will remain confidential and will be disclosed only with your permission or as required by law. No data collected in this study will be released to any other party for any reason. All data from this study will be stored in a file maintained on a password-protected computer owned by Purdue and utilized only by the researchers. Data will be stored on secure computers in Wang Hall at Purdue University West Lafayette and in the research group's secure data depot online. The study results will be used for reports or publications, but will only be shared in aggregate form, i.e., means or standard deviations. Video/audio data will be kept confidential and used for transcription and coding purposes. Video/audio data will be destroyed at the end of the project.

What are my rights if I take part in this study?

Your participation in this study is voluntary. You may choose not to participate or, if you agree to participate, you can withdraw your participation at any time without penalty.

Who can I contact if I have questions about the study?

If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact Prof. Jennifer DeBoer, deboerj@purdue.edu, Prof. Ed Berger, bergere@purdue.edu, or Prof. Jeff Rhoads, jfrhoads@purdue.edu. If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu)or write to:

Human Research Protection Program - Purdue University Ernest C.
Young Hall, Room 1032
155 S. Grant St.,
West Lafayette, IN 47907-2114

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am at least 18 years old and am prepared to participate in the research study described above. I will be offered a copy of this consent form after I sign it.

______________________________________ ______________________

Participant's Signature Date

____

Participant's Name

______________________________________ ______________________

Researcher's Signature Date

APPENDIX B. INTERVIEW CONSENT FORM

Purdue IRB Protocol #: 1507016258 - Expires on: 08-JUN-2018

RESEARCH PARTICIPANT CONSENT FORM

Understanding and Supporting Mechanical Engineering Undergraduate Student and Faculty Engagement with an Active, Blended and Collaborative (ABC) Learning Environment

PI Jennifer DeBoer
Engineering Education
Purdue University

What is the purpose of this study?

You are asked to participate in a research study conducted by Profs. Jennifer DeBoer, Ed Berger, Jeff Rhoads, and Chuck Krousgrill of the Engineering Education and Mechanical Engineering Departments of Purdue University. You were selected as a participant in this study because you are registered in **Course Number** this semester. You should read the information below before deciding whether or not to participate.

The purpose of this study is to understand the use and possible impact of the *Freeform* class structure. This information may help to improve future **Course Number/Name** classes at Purdue and elsewhere.

What will I do if I choose to be in this study?

- This interview/problem solving session is voluntary. You have the right not to answer any question, and to stop the interview at any time or for any reason. We expect that the interview/observation will take 20-30 minutes.
- While you work, eye gaze software will record what components of the online resources you are looking at. You will be asked to use a Livescribe pen for any written notes you choose to use, which will record what you write for easier analysis.
- If you are participating in the observation, the problem will be of equivalent difficulty to a homework problem at the point in the semester during which the observation takes place. Your success on the problem will not have any bearing on your grade.
- Interviews and observations will take place in a restricted use room in the WANG 3500 Engineering Education suite.

How long will I be in the study?

You will be a participant in the study for the duration **Course Number**. We will analyze your information after the course is over, and your participation in the study will have no effect on your grade in **Course Number**.
The interview will take approximately 20-30 minutes. The problem solving session will take no more than 30 minutes. You will complete no more than 2 interview/problem solving sessions per course.

What are the possible risks or discomforts?

Minimal possible risks or discomfort. Breach of confidentiality is the only risk and the safeguards used to minimize this risk can be found in the confidentiality section. Participation in this study will have no relationship or consequences for your grade in **Course Number**.

Are there any potential benefits?

There are no direct benefits to the participant. Your participation may provide you the chance to share your experiences with the new type of learning environment in **Course Number**. Your valuable perspective may help to improve this environment for future students.

Purdue IRB Protocol #: 1507016258 - Expires on: 08-JUN-2018

Will I receive payment or other incentive?

You will be compensated $20 for completion of the interview/observation. In addition, the information that is learned will ultimately benefit other faculty and students who participate in a Freeform structure course.

Will information about me and my participation be kept confidential?

The project's research records may be reviewed by the National Science Foundation and by departments at Purdue University responsible for regulatory and research oversight.

- Information you tell us will be confidential, and results will be published as anonymized and de-identified, showing overall results in the aggregate.
- We would like to record this interview in a digital file so that we can use it for reference while proceeding with this study. We will not record this interview without your permission. If you do grant permission for this conversation to be recorded digitally, you have the right to revoke recording permission and/or end the interview at any time. The interview will be transcribed by a service used by Purdue's Engineering Education department.

Any information that is obtained in connection with these data collections and that may be identified with you will remain confidential and will be disclosed only with your permission or as required by law. No data collected in this study will be released to any other party for any reason. All data from this study will be stored in a file maintained on a password-protected computer owned by Purdue and utilized only by the researchers. Data will be stored on secure computers in Wang Hall at Purdue University West Lafayette and in the research group's secure data depot online. The study results will be used for reports or publications, but will only be shared in aggregate form, i.e., means or standard deviations. Video/audio data will be kept confidential and used for transcription and coding purposes. Video/audio data will be destroyed at the end of the project.

What are my rights if I take part in this study?

Your participation in this study is voluntary. You may choose not to participate or, if you agree to participate, you can withdraw your participation at any time without penalty or loss of benefits to which you are otherwise entitled.

Who can I contact if I have questions about the study?

If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact Prof. Jennifer DeBoer, deboerj@purdue.edu, Prof. Ed Berger, bergere@purdue.edu, or Prof. Jeff Rhoads, jfrhoads@purdue.edu.

If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu)or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.,
West Lafayette, IN 47907-2114

Documentation of Informed Consent

Purdue IRB Protocol #: 1507016258 - Expires on: 08-JUN-2018

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am prepared to participate in the research study described above. I will be offered a copy of this consent form after I sign it.

(*Please check all that apply*)

[] I give permission for this interview to be recorded digitally and transcribed

[] I give permission for my responses (anonymized/aggregated) included in this study

____ _________________________

Participant's Signature Date

____

Participant's Name

____ ___________________________

Researcher's Signature Date

APPENDIX C. RESOURCE USAGE SURVEY

Start of Block: Percentage of Time Spent Studying Alone

SAQ This survey begins with questions about your <u>ME 274 work habits</u>, and we are specifically interested in your relationships to your peers. In the following questions:

ALONE means you are working by yourself. **IN A GROUP** means you are working with <u>at least one other person</u>. This work time could be planned ahead of time ("Let's meet in the ME building at 7 pm.") <u>**OR**</u> it could be unplanned ("I see that you are working on ME 274, would you like to work together on it?").

In this survey, we use the terms <u>**WORK**</u> **or** <u>**WORKING**</u> to indicate any activity in which you engage that contributes to your academic success in ME 274.

Please estimate the percentage of your <u>out of class</u> work time for dynamics that you spend doing the following:

Percentage of time I spend working on dynamics **ALONE**. (1)	

End of Block: Percentage of Time Spent Studying Alone

Start of Block: Study Group Questions

SG1 When you work on dynamics **IN A GROUP**, what is the typical size of the group, including yourself?

Number of people typically present when I work **IN A GROUP**. (1)	

SG2 In a typical week of working on dynamics:

About how many hours do you work outside of class **ALONE**? (1)	
About how many hours do you work outside of class **IN A GROUP**? (2)	

SG3 During the week before a dynamics exam:

About how many hours do you work outside of class **ALONE**? (1)	
About how many hours do you work outside of class **IN A GROUP**? (2)	

SG4 In a typical week of working on dynamics:

About how many work sessions do you engage in **ALONE**? (1)	
About how many work sessions do you engage in **IN A GROUP**? (2)	

SG5 When you work **IN A GROUP**, indicate how frequently you meet in various locations:

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
In an academic space, like a vacant classroom or the library. (1)	○	○	○	○	○
In a residential space, like an apartment, dorm room, or fraternity/sorority house. (2)	○	○	○	○	○
In another public location, like a coffee shop or the Union. (3)	○	○	○	○	○

SG6 When you work **ALONE**, indicate how frequently you work in various locations:

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
In an academic space, like a vacant classroom or the library. (1)	○	○	○	○	○
In a residential space, like an apartment, dorm room, or fraternity/sorority house. (2)	○	○	○	○	○
In another public location, like a coffee shop or the Union. (3)	○	○	○	○	○

SG7 When you work **ALONE** or **IN A GROUP**, indicate how frequently you work during various times of the day:

	ALONE					**IN A GROUP**				
	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
During the day, before 5 pm. (1)	○	○	○	○	○	○	○	○	○	○
During the evening, between 5-10 pm. (2)	○	○	○	○	○	○	○	○	○	○
Late at night, after 10 pm. (3)	○	○	○	○	○	○	○	○	○	○

End of Block: Study Group Questions

Start of Block: Blog Questions

B1 This question asks about your **BLOG USAGE**.

The course blog has a comment section that allows students to interact with each other students, ask questions, answer questions, and share information. Of the following choices, which are reasons why you ***WOULD NOT*** routinely use the blog ***COMMENT*** features? (Check all that apply.)

▢ I prefer to interact with my peers face to face. (1)

▢ I prefer to ask the instructor for help in person. (2)

▢ The lecture book is a good source of information for me. (3)

▢ The online solution videos (both lecture examples and homework) generally answer my questions for me. (4)

▢ I don't like the 'public' nature of the blog comment threads. (5)

▢ I don't think I have anything to contribute. (6)

▢ I didn't have time to be active on the blog. (7)

▢ Not enough of my peers use the blog for it to be useful. (9)

▢ Other (8) ____

End of Block: Blog Questions

Start of Block: Video Usage Questions

VU1 This next set of questions asks about your **USAGE OF THE LECTURE EXAMPLE AND HOMEWORK SOLUTION VIDEOS.**

When you watch lecture example or homework solution videos, which choice best describes your usual approach?

- ◯ I usually watch the entire video, from beginning to end. (1)
- ◯ I usually search for and watch specific parts of the video that I think will be useful or informative. (2)
- ◯ My strategy in watching the video usually depends upon what problem I'm trying to solve. (3)

VU2 What is the ideal length of a single video (in minutes), to maintain your attention?

Video length (minutes) (1)	

VU3 About how many videos (lecture example or homework solution) have you watched this semester (in part or in full)?

Approximate number of videos watched this semester (1)	

VU4 How frequently are you doing the following when you watch lecture example or homework solution videos?

	Never (1)	Rarely (2)	About half the time (3)	Most of the time (4)	Always (5)
I am alone. (1)	○	○	○	○	○
I am with other people. (2)	○	○	○	○	○
I am doing homework. (3)	○	○	○	○	○
I am preparing for class. (9)	○	○	○	○	○
I am preparing for an exam. (4)	○	○	○	○	○
I am checking the solution to my homework after it has been graded and returned to me. (8)	○	○	○	○	○
I watch on my computer or tablet device. (5)	○	○	○	○	○
I watch on my phone. (6)	○	○	○	○	○

VU5 Of the following choices, which one best describes the way you usually use the solution videos to support your problem solving.

- ◯ I watch the video in its entirety before I attempt to solve the problem. (1)
- ◯ I write the solution to the problem in the video, as the video is playing. (2)
- ◯ I watch part of the video, pause it, and try to complete the next step on my own. Then I check my work with the video. (3)
- ◯ I solve the entire problem on my own, then use the video to check my work. (4)
- ◯ Other (5) ____

End of Block: Video Usage Questions

Start of Block: Help Seeking Behaviors

HS1 This section of the survey focuses on **how you access academic help when you need it in dynamics**. Please identify **how FREQUENTLY you use each of the following resources for help** in dynamics.

	At least once per day (1)	3-6 times per week (2)	1-2 times per week (3)	1-3 times per month (4)	1-3 times per semester (5)	Never (99)
my peers in the class (1)	○	○	○	○	○	○
the course lecturebook (2)	○	○	○	○	○	○
the lecture example and homework solution videos (3)	○	○	○	○	○	○
the course blog (4)	○	○	○	○	○	○
the instructor, by asking questions in class (5)	○	○	○	○	○	○
the instructor, during office hours (6)	○	○	○	○	○	○
online resources not accessed from the course blog (ex: online lectures or videos not associated with the course) (7)	○	○	○	○	○	○
other students I know who are not currently enrolled in the class (8)	○	○	○	○	○	○
the TAs in the mechanics tutorial room (9)	○	○	○	○	○	○

HS2 For the resources you used **LEAST FREQUENTLY**, for which of the following reason(s) do you access those resources so infrequently? Check all that apply.

- ▢ The resources are rarely helpful. (1)
- ▢ The resources are too hard to access (example: too far away from where I usually study). (2)
- ▢ The resources are not available when I need it (example: office hours conflict with my course schedule). (3)
- ▢ For another reason (please specify in the text box below). (4) ____

HS3 Please identify **HOW USEFUL** the following resources are for the dynamics course.

	Completely useless (1)	Somewhat useless (2)	Neither useless nor useful (3)	Somewhat useful (4)	Very useful (5)	No opinion (99)
my peers in the class (1)	○	○	○	○	○	○
the course lecturebook (2)	○	○	○	○	○	○
the lecture example and homework solution videos (3)	○	○	○	○	○	○
the course blog (4)	○	○	○	○	○	○
the instructor, by asking questions in class (5)	○	○	○	○	○	○
the instructor, during office hours (6)	○	○	○	○	○	○
online resources not accessed from the course blog (ex: online lectures or videos not associated with the course) (7)	○	○	○	○	○	○
other students I know who are not currently enrolled in the class (8)	○	○	○	○	○	○
the TAs in the mechanics tutorial room (9)	○	○	○	○	○	○

End of Block: Help Seeking Behaviors

Start of Block: Experience and Comparison Questions

EC1 This block of questions asks you to **COMPARE YOUR EXPERIENCE IN DYNAMICS THIS SEMESTER** with your experience in **OTHER COURSES**.

This question asks you to compare your experience in dynamics this semester with your experiences in the other courses in which you were enrolled this semester. Choose the option that reflects your level agreement with each statement.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
In dynamics, I felt a stronger connection to my peers. (1)	○	○	○	○	○
The dynamics class had more online resources available to help me. (2)	○	○	○	○	○
Dynamics was taught in a way that aligns more closely with my preferred way of learning. (3)	○	○	○	○	○
The dynamics course was more organized and structured. (4)	○	○	○	○	○
I felt a strong connection to my dynamics instructor. (6)	○	○	○	○	○
I found it harder to be successful in dynamics than it was to be successful in my other courses. (5)	○	○	○	○	○

EC2 These questions ask you to compare your experience in dynamics this semester with your experiences in the other courses in which you have enrolled that **PROVIDED SIGNIFICANT ONLINE CONTENT** as part of the course. Choose the option that reflects your level of agreement with each statement. (If you have **NEVER TAKEN** another course with significant online content, you can skip this question.)

	strongly disagree (1)	somewhat disagree (2)	neither agree nor disagree (3)	somewhat agree (4)	strongly agree (5)
In dynamics, I felt a stronger connection to my peers. (1)	○	○	○	○	○
The dynamics class had more online resources available to help me. (2)	○	○	○	○	○
Dynamics was taught in a way that aligns more closely with my preferred way of learning. (3)	○	○	○	○	○
The dynamics course was more organized and structured. (4)	○	○	○	○	○
I felt a stronger connection to my dynamics instructor. (6)	○	○	○	○	○
I found it harder to be successful in dynamics than it was to be successful in those other courses. (5)	○	○	○	○	○

EC3 In what ways did the actual experience of the dynamics course meet, not meet, or exceed the expectations you had at the start of the semester?

○ The dynamics course **MET** my expectations in the following ways: (1)

____

○ The dynamics course **DID NOT MEET** my expectations in the following ways: (2)

____

○ The dynamics course **EXCEEDED** my expectations in the following ways: (3)

____

EC4 For each category below, please give us your recommendations about how to improve your experience in dynamics. For each item, tell us whether we should increase, decrease, or keep that item about the same.

	increase this (1)	keep this about the same (2)	decrease this (3)
quantity of homework (1)	○	○	○
difficulty of homework (2)	○	○	○
number of exams (3)	○	○	○
difficulty of exams (4)	○	○	○
instructor office hours (5)	○	○	○
number of lecture example and homework solution videos (6)	○	○	○
amount of discussion activity on the blog (7)	○	○	○
amount of active learning in class (8)	○	○	○
amount of lecturing in class (9)	○	○	○
opportunities to formally work in teams (ex.: a team-based course project) (10)	○	○	○
other suggestion (please describe in the box) (11)	○	○	○

EC5 What is your single most important suggestion about how to improve the student experience in dynamics?

____

EC6 Knowing what you know now, what advice would you give to other students about how to be successful in dynamics?

____

End of Block: Experience and Comparison Questions

APPENDIX D. MOTIVATION AND TASK VALUE SURVEY

Freeform   Updated: Fall 2017

Student Survey

Please use a #2 pencil to fill in your answers on the bubble sheet.

Name (please print): ___FILL IN ON BUBBLE SHEET (FIRST AND LAST)___

Student Identification Number: ___FILL IN ON BUBBLE SHEET (must be 10 digits)___

Please indicate how true each of the statements is for you by **filling in the circle on the bubble sheet** corresponding to your response. The response 1 = "not at all true of me", and 7 = "very true of me".

		Not at All True of Me ↓ 1	2	3	4	5	6	Very True of Me ↓ 7
1.	I will use the information I learn in ME 274 in the future.	1	2	3	4	5	6	7
2.	When I take a test, I think about how poorly I am doing compared with other students.	1	2	3	4	5	6	7
3.	I usually study in a place where I can concentrate on my course work.	1	2	3	4	5	6	7
4.	I have an uneasy, upset feeling when I take an exam.	1	2	3	4	5	6	7
5.	I feel my heart beating fast when I take an exam.	1	2	3	4	5	6	7
6.	I am confident I can do an excellent job on the assignments and tests in ME 274.	1	2	3	4	5	6	7
7.	I make good use of my study time for my courses.	1	2	3	4	5	6	7
8.	I expect to do well in ME 274.	1	2	3	4	5	6	7
9.	I find it hard to stick to a study schedule.	1	2	3	4	5	6	7
10.	The grade I get in ME 274 will not be important for my future success.	1	2	3	4	5	6	7
11.	I am certain I can master the skills being taught in ME 274.	1	2	3	4	5	6	7
12.	I have a regular place set aside for studying.	1	2	3	4	5	6	7
13.	Considering all the aspects of the course – the difficulty of this course, the teacher, and my skills – I think I will do well in ME 274.	1	2	3	4	5	6	7

		Not at All True of Me ↓						Very True of Me ↓
		1	2	3	4	5	6	7
14.	I make sure I keep up with the weekly readings and assignments for my courses.	1	2	3	4	5	6	7
15.	I attend class regularly.	1	2	3	4	5	6	7
16.	When I take tests, I think of the consequences of failing.	1	2	3	4	5	6	7
17.	The grade I get in ME 274 will affect my future.	1	2	3	4	5	6	7
18.	What I learn in ME 274 will be important for my future occupational success.	1	2	3	4	5	6	7
19.	When I take a test, I think about items on other parts of the test I can't answer.	1	2	3	4	5	6	7
20.	I will use the information I learn in ME 274 in other classes I will take in the future.	1	2	3	4	5	6	7
21.	I believe I will receive an excellent grade in ME 274.	1	2	3	4	5	6	7
22.	I must pass ME 274 in order to reach my academic goals.	1	2	3	4	5	6	7
23.	I am certain I can understand the most difficult material presented in ME 274.	1	2	3	4	5	6	7
24.	I often find that I don't spend very much time on my courses because of other activities.	1	2	3	4	5	6	7
25.	The grade I get in ME 274 will not affect my ability to continue on with my education.	1	2	3	4	5	6	7
26.	I am confident I can understand the most complex material presented by the instructor in ME 274.	1	2	3	4	5	6	7
27.	I rarely find time to review my notes or readings before exams.	1	2	3	4	5	6	7
28.	I will not use what I learn in ME 274.	1	2	3	4	5	6	7
29.	I am confident I can understand the basic concepts taught in ME 274.	1	2	3	4	5	6	7

[CONTINUE SURVEY ON NEXT PAGE]

The next set of questions asks about your views on educational technology for your own learning. Here, educational technology can refer to any online or mobile tools and resources that are available to you to support your learning (such as a course blog, in-class clickers, or mobile apps).

To what extent do you agree with the following statements? **Fill in the circle on the bubble sheet** corresponding to your response. The response 1 = "strongly disagree", and 7 = "strongly agree".

Educational technology...

		Strongly disagree ↓ 1	2	3	4	5	6	Strongly agree ↓ 7
30.	...helps me develop skills in planning and time management.	1	2	3	4	5	6	7
31.	...improves my academic performance.	1	2	3	4	5	6	7
32.	...helps me to consolidate and process information more effectively.	1	2	3	4	5	6	7
33.	...only creates organizational problems for me in my classes.	1	2	3	4	5	6	7
34.	...helps me work at a level appropriate to my learning needs.	1	2	3	4	5	6	7
35.	...results in poorer calculation and estimation skills for me.	1	2	3	4	5	6	7
36.	...helps me learn to collaborate with other students.	1	2	3	4	5	6	7
37.	...limits the amount of personal communication I have with other students.	1	2	3	4	5	6	7
38.	...only encourages me to copy material from published internet sources.	1	2	3	4	5	6	7
39.	...enables me to access better sources of information.	1	2	3	4	5	6	7
40.	...impedes my learning of concepts better done with real objects.	1	2	3	4	5	6	7
41.	...enables me to communicate more effectively with others.	1	2	3	4	5	6	7
42.	...helps me develop greater interest in learning.	1	2	3	4	5	6	7
43.	...only distracts me from learning.	1	2	3	4	5	6	7
44.	...results in poorer writing skills for me.	1	2	3	4	5	6	7

APPENDIX E. IJEE COPYRIGHT PERMISSION

Gmail **Nick Stites <nstites.purdue@gmail.com>**

Revised Manuscript: Bias in the Assessment, Validation Model, or Both? ... [A Validation and Differential Item Functioning (DIF) Study of an Abbreviated Dynamics Concept Inventory]

---

ijee <ijee.editor@gmail.com> Fri, Nov 16, 2018 at 12:31 PM
To: nstites@purdue.edu

Good afternoon Nick,

The manuscript will be scheduled for publication in an upcoming issue of the IJEE, most probably issue 35-2 (the March/April issue).

Please feel free to use the contents of the paper in your thesis.

I think this e-mail will sufficient to serve as approval of using the contents. The IJEE does not have a standard form for this purpose.
However, If you need a more formal letter I can prepare one for you you one.

All the best,

Ahmad Ibrahim

On Fri, Nov 16, 2018 at 10:45 AM Nick Stites <nstites@purdue.edu> wrote:

> Dear Dr. Ahmad Ibrahim,
>
> Recently, I sent you our revisions to the accepted manuscript entitled "A Validation and Differential Item Functioning (DIF) Study of an Abbreviated Dynamics Concept Inventory." I am including this manuscript in my PhD thesis, which I am defending on Nov. 26 and depositing by Dec. 7. **Because this manuscript is accepted, would you please send me written approval for the use of this content in my thesis?** Purdue University does not have a specific form for this copyright release; do you have a standard form/letter you send to authors in similar circumstances? The work is fully cited directly under the chapter title in my thesis.
>
> Also, could you provide an update on the review status of our revisions?
>
> Thank you for your time and help,
>
> Nick A. Stites
>
> On Fri, Oct 26, 2018 at 10:11 PM ijee <ijee.editor@gmail.com> wrote:
>
>> Thank you for your e-mail and the attached documents.
>>
>> On Fri, Oct 26, 2018 at 11:16 PM Nick Stites <nstites@purdue.edu> wrote:
>>
>>> Dear Dr. Ahmad Ibrahim,
>>>
>>> Thank you for the time and effort you and the IJEE team have given to reviewing our manuscript now entitled "A Validation and Differential Item Functioning (DIF) Study of an Abbreviated Dynamics Concept Inventory." We are excited that IJEE has accepted this manuscript, and we have completed the requested revisions.
>>>
>>> Our revised manuscript and our responses to the reviewers' comments are attached. Please let me know if you have any questions or concerns regarding our revised version. I look forward to working with you and the IJEE team to advance this manuscript toward publication.
>>>
>>> Sincerely,

APPENDIX F. LIST OF ACRONYMS USED IN CHAPTER 2

Acronym	Definition	Description
aDCI	Abbreviated Dynamics Concept Inventory	Selection of 12 items from the DCI
CFA	Confirmatory factor analysis	Method for testing latent structures
CFI	Comparative fit index	Goodness of fit statistic
CI	Concept inventory	Usually multiple-choice tests that require little or no calculations
DCI	Dynamics Concept Inventory	29-item dynamics concept inventory
df	Degrees of freedom	Measure of how much data is available relative to how many model parameters are being estimated
DIF	Differential item functioning	Scenario of an item functioning differently for distinct groups
FCI	Force Concept Inventory	Physic concept inventory
IRT	Item response theory	Method of modeling latent ability and item characteristics
MG-CFA	Multiple-group confirmatory factor analysis	Method for testing the invariance of a measurement model across multiple groups
RMSEA	Root-mean square error of approximation	Goodness of fit statistic
χ^2	Chi-squared test statistic	Goodness of fit statistic
3PL	Three parameter model	IRT method that models an items difficulty, discrimination, and guessing parameter

APPENDIX G. ADDITIONAL METHOD DETAILS FOR CHAPTER 3

Participants

Approximately 500 students enrolled in Dynamics each year, with the most students (over 350) enrolling during the spring semester. The total number of survey responses was 581, comprised of 95, 36, 139, 83, and 228 responses from the Spring 2016, Fall 2016, Spring 2017, Fall 2017, and Spring 2018 semesters, respectively. From Spring 2016-Spring 2017, no incentive was given for completing the survey. For the Fall 2017 and Spring 2018 semesters, additional questions from a partner organization were appended to the original survey, and ten points of extra credit toward the student's homework grade (which amounted to less than 0.45% of extra credit toward a student's overall grade in Dynamics) were given to anyone who completed the survey.

Regarding the qualitative data, students were offered a $20 gift card for participating in an interview. Participants were recruited through email, using a stratified sampling strategy based on prior GPA, section (instructor), and international status to capture the experiences of different student groups. These stratifications reflect the fact that the student interviews were used to collect data for this study and many others not discussed here (e.g., how the students' experiences in Dynamics differed across instructors with varying levels of experience teaching the course). Regarding the international-status stratification, we expected that a student's experiences in the class may vary more according to a student's international status than according to their gender or major because of language and cultural differences.

Our goal was to interview at least two students from each of the stratified student groups. If the number of participants was low in a given stratified group, then up to two follow-up recruitment emails were sent to that group. Over-participation from a stratified group was allowed as no volunteer was denied an interview.

A total of 53 interviews with students who also completed the end-of-semester survey were completed between Spring 2016-Fall 2017 (there were none conducted during Spring 2018), but this study only utilized 44 of those interviews. We used a subsample of the interviews because we wanted to better understand the archetypical resource-usage behaviors of the students. Therefore, if a student's usage pattern did not meet a threshold for alignment with one of the most-common patterns of resource usage, as further explained in the Data Analysis section, the student's interview data was not used in the qualitative analysis.

Data Analysis

Interviewee Selection

The cluster analysis results were used to subsample the student interviewees so that we only analyzed the interview transcripts of students who exhibited resource-usage behavior that aligned well with one of the archetypical resource-usage patterns. We used a measure of cluster-membership uncertainty to do this, which is calculated as unity minus the largest cluster-membership probability—see the Data Analysis section of the main paper for more details. The qualitative analysis only included students with cluster-membership uncertainties of less than 0.30. When considering the entire sample, approximately 84% of the students had an uncertainty less than 0.30 (the mean uncertainty was 0.11, and the median was 0.02). We also considered a lower uncertainty threshold of 0.10 (~68% of the sample had an uncertainty of less than 0.10), but the number of students who completed an interview and had uncertainties less than 0.90 was ten students fewer than a threshold of 0.30. Of these ten students, five were women and four were not mechanical-engineering majors; all were domestic students. Given that the cluster analysis is largely driven by the patterns of resource-usage of the majority—White, domestic men majoring in mechanical engineering—we decided to use the higher uncertainty threshold of 0.30 in order to improve the diversity of the interviews included in the qualitative analysis.

Qualitative analysis.

We used thematic analysis to find themes from the students' interview transcripts in their descriptions and explanations of their resource-usage. We used a thematic-analysis process based on the recommendations of Braun and Clarke (2006). The coding happened in two phases. First, to categorize the content of the interview, we read through each interview and coded the content related to each of the resources listed in Table 1 of the main paper. For this categorization of content, we used a coding scheme developed by Kandakatla et al. (in review). This coding was completed mostly by one undergraduate research assistant, and two inter-rater reliability checks with one of the authors ensured coding consistency. Then, for the second phase of coding, we reread the transcripts and then, using the codes from the first phase, extracted the resource-related content that corresponded to the unique resource-usage characteristics of each cluster. This extracted content for students in a given cluster was coded for interesting features. These initial codes were then grouped according to potential themes, and the transcripts were reread to ensure that these potential themes accurately represented the students' thoughts and words. The final themes were summarized for each cluster. Lastly, the themes were viewed through the expectancy-value conceptual framework for resource selection to determine what factors of the expectancy-value model seemed to influence the resource usage of the students in each cluster.