

Spring 2015

# Difficulty as a concept inventory design consideration: An exploratory study of the concept assessment tool for statics (CATS)

Dana L. Denick  
*Purdue University*

Follow this and additional works at: [https://docs.lib.purdue.edu/open\\_access\\_dissertations](https://docs.lib.purdue.edu/open_access_dissertations)

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Engineering Education Commons](#)

---

## Recommended Citation

Denick, Dana L., "Difficulty as a concept inventory design consideration: An exploratory study of the concept assessment tool for statics (CATS)" (2015). *Open Access Dissertations*. 448.  
[https://docs.lib.purdue.edu/open\\_access\\_dissertations/448](https://docs.lib.purdue.edu/open_access_dissertations/448)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

**PURDUE UNIVERSITY  
GRADUATE SCHOOL  
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Dana L. Denick

Entitled

Difficulty as a Concept Inventory Design Consideration: An Exploratory Study of the Concept Assessment Tool for Statics (CATS)

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Ruth Streveler

Chair

Monica Cardella

Senay Purzer

Paul Steif

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Ruth Streveler

Approved by: Ruth Streveler

Head of the Departmental Graduate Program

4/9/2015

Date



DIFFICULTY AS A CONCEPT INVENTORY DESIGN CONSIDERATION: AN  
EXPLORATORY STUDY OF THE CONCEPT ASSESSMENT TOOL FOR STATICS  
(CATS)

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Dana L. Denick

In Partial Fulfillment of the  
Requirements for the Degree

of

Doctor of Philosophy

May 2015

Purdue University

West Lafayette, Indiana

This dissertation is dedicated to the memory of my father

David J. Denick (1942-2012)

Your commitment to creativity and education continues to inspire me

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Ruth Streveler for her guidance and support throughout my graduate studies at Purdue. I can't imagine experiencing such a valuable, respectful and caring advising relationship with anyone else. I would also like to thank my committee, Dr. Paul Steif, Dr. Senay Purzer, and Dr. Monica Cardella for their counsel and encouragement while working on my dissertation.

I am grateful for my writing group and extended research family, especially Nicole Pitterson, Farrah Fayyaz, Natasha Perova, and Aidsa Santiago-Román. Each of you has helped me strengthen my work on this study and grow as a researcher. A special thanks to Emily Dringenberg, whose enthusiasm and friendship kept me going.

I would not have been able to complete this dissertation without the support of my family. Most of all, I want to thank my mother for always believing I could get it done. You're my source of inspiration and perseverance; thank you for everything.

This work was supported by the National Science Foundation under grant DRL 0918531. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author do not necessarily reflect the views of the National Science Foundation.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
ABSTRACT .....	xii
CHAPTER 1. INTRODUCTION .....	1
1.1 Background Information .....	1
1.2 Problem Statement .....	1
1.3 Statement of Purpose.....	2
1.4 Research Question.....	3
1.5 Overview of Methodology .....	4
1.6 Rationale and Significance.....	4
1.7 Organization of Dissertation .....	6
CHAPTER 2. LITERATURE REVIEW .....	8
2.1 Introduction .....	8
2.1.1 Development of Engineering Concept Inventories.....	8
2.1.2 The Assessment Triangle.....	10
2.2 Expanding upon the Cognitive Element of the Assessment Triangle.....	13
2.2.1 Nature of Conceptual Understanding .....	13
2.2.2 Statics Concepts and Conceptual Framework .....	16
2.3 Expanding upon the Observation Element of the Assessment Triangle .....	19
2.3.1 Concept Inventory Item Design.....	20
2.3.2 CATS Item Design .....	21
2.3.3 Concept Inventory Development Process.....	22
2.3.4 Development of CATS .....	23

	Page
2.4 Expanding upon the Interpretation Element of the Assessment Triangle .....	26
2.4.1 Psychometric Evaluation .....	27
2.4.2 Psychometric Evaluation of CATS.....	29
2.4.3 The Interpretation Vertex as a Validity Argument .....	29
2.4.4 A Comprehensive Approach to Validity for Concept Inventories .....	30
2.5 Addressing Gaps in Knowledge .....	33
2.5.1 Cognitive Load Theory.....	33
2.5.2 Evidence-Centered Design .....	34
2.6 Summary .....	36
2.6.1 Interconnectedness of the Assessment Triangle .....	36
CHAPTER 3. METHOD.....	38
3.1 Introduction .....	38
3.2 Research Design .....	38
3.2.1 Mixing Research Paradigms .....	40
3.3 Data Sources.....	41
3.3.1 CATS .....	42
3.3.2 Source Material for Content Analysis .....	42
3.3.3 Interview Participants .....	45
3.4 Data Collection.....	47
3.4.1 Content Analysis Data Collection .....	47
3.4.2 Think-aloud Interviews.....	47
3.4.3 Interview Materials .....	47
3.4.4 Interview Protocol .....	48
3.4.5 Use of Existent Data .....	48
3.5 Method of Analysis .....	49
3.5.1 CATS Item Grouping .....	49
3.5.2 Content Analysis.....	51
3.5.3 Thematic Analysis .....	56
3.6 Theme Definitions.....	60



	Page
3.6.1 Direction of Problem Solving.....	61
3.6.2 Distractor Attractiveness .....	66
3.7 Quality Considerations .....	67
CHAPTER 4. FINDINGS .....	69
4.1 Introduction .....	69
4.2 Direction of Problem Solving .....	69
4.2.1 Direction of Problem Solving at the Item-Level .....	70
4.2.2 Direction of Problem Solving at the Student Response-level .....	71
4.3 Distractor Attractiveness .....	74
4.3.1 Distractor Attractiveness of Less Difficult CATS Items .....	75
4.3.2 Distractor Attractiveness of More Difficult CATS Items.....	81
4.3.3 Distractor Attractiveness of Items with Difficult Cognitive Attributes.....	86
4.3.4 Summary of Distractor Reasoning.....	91
4.3.5 Distractor Attractiveness by Item Group.....	93
4.3.6 Distractor Attractiveness by Student .....	94
4.4 Summary .....	96
CHAPTER 5. CONCLUSION.....	98
5.1 Introduction .....	98
5.2 Discussion of Findings from Thematic Analysis .....	99
5.2.1 Discussion of Findings from Direction of Problem Solving Theme .....	100
5.2.2 Discussion of Findings from Distractor Attractiveness Theme.....	101
5.3 Conclusions .....	105
5.3.1 Theory of Difficulty for CATS.....	106
5.3.2 Relation to Cognitive Load Theory .....	107
5.3.3 Interrelationship among Item Design, Student Reasoning and Difficulty...	109
5.3.4 Relation to Evidence-Centered Design Framework .....	110
5.4 Implications and Recommendations .....	111
5.4.1 Implications for CATS as an Instructional Tool.....	111
5.4.2 Implications for Concept Inventory Development .....	113

	Page
5.4.3 Future Research Opportunities .....	114
REFERENCES .....	117
APPENDICES	
Appendix A IRB Exemption .....	129
Appendix B Interview Protocol.....	131
Appendix C Content Analysis.....	151
VITA.....	175

## LIST OF TABLES

Table	Page
Table 2.1 Selection of Concept Inventories Developed for use in Engineering Education .....	10
Table 2.2 Clusters of Concepts for the Conceptual Framework of Statics (from Steif & Dantzler, 2005, p. 363) .....	18
Table 2.3 Conceptual Errors in Statics (from Steif & Dantzler, 2005, p. 364).....	19
Table 2.4 Concepts addressed on each item (from Steif & Hansen, 2007, p. 206) .....	22
Table 2.5 Relation of Data Collection to Validity Components as part of a Comprehensive Approach to Validity (adapted from (Pellegrino et al., 2013)).....	32
Table 3.1 CATS Design and Validity Study Literature .....	43
Table 3.2 CATS Item Difficulty as Determined through Classical Test Theory, and 1- and 2-Parameter Item Response Theory .....	50
Table 3.3 Problematic Cognitive Attributes as Determined by the Fusion Model.....	51
Table 3.4 Case Groups by Difficulty .....	51
Table 4.1 Item-level Coding within Direction of Problem Solving Theme .....	70
Table 4.2 Student Response-level Coding within Direction of Problem Solving Theme, Less Difficult CATS Items .....	72
Table 4.3 Student Response-level Coding within Direction of Problem Solving Theme, More Difficult CATS Items .....	73

Table	Page
Table 4.4 Student Response-level Coding within Direction of Problem Solving Theme, CATS Items with Difficult Cognitive Attributes.....	73
Table 4.5 Distractor Attractiveness Coding for CATS Items 1 & 3.....	76
Table 4.6 Distractor Attractiveness Coding for CATS Items 14 & 15.....	78
Table 4.7 Distractor Attractiveness Coding for CATS Item 19.....	80
Table 4.8 Distractor Attractiveness Coding for CATS Items 4 & 5.....	82
Table 4.9 Distractor Attractiveness Coding for CATS Items 7 & 8.....	84
Table 4.10 Distractor Attractiveness Coding for CATS Items 22 & 23.....	86
Table 4.11 Distractor Attractiveness Coding for CATS Item 17.....	87
Table 4.12 Distractor Attractiveness Coding for CATS Item 18.....	89
Table 4.13 Distractor Attractiveness Coding for CATS Item 27.....	90
Table 4.14 Summary of Expected and Observed Problematic Reasoning Coded to Item Distractors.....	92
Table 4.15 Distractor Attractiveness by Item Group, Codes are Indicated as Misconception (C), Misapplication (A), and Partial (P).....	93
Table 4.16 Distractor Attractiveness by Student, Codes are indicated as Misconception (C), Misapplication (A), Partial (P), and Other (O).....	95

## LIST OF FIGURES

Figure	Page
Figure 1.1 The Assessment Triangle (Pellegrino et al., 2001).....	5
Figure 2.1 The Assessment Triangle (Pellegrino et al., 2001) including (a) elements of the assessment triangle and (b) as applied to CATS.....	11
Figure 2.2 Concept Inventory Item Components.....	20
Figure 2.3 Sample CATS Item.....	21
Figure 2.4 Problematic Concept Inventory Development Process Flowchart, (Santiago-Román, 2009).....	25
Figure 2.5 Revised Concept Inventory Development Process Flowchart, modified from (Santiago-Román, 2009).....	26
Figure 2.6 Proposed Data Sources and Collection Strategies as part of a Comprehensive Approach to Validity (adapted from (Pellegrino et al., 2013)).....	32
Figure 2.7. Simplified Representation of the Conceptual Assessment Framework of Evidence-Centered Design (Pellegrino et al., 2014).....	35
Figure 3.1 Research Design Showing the Qualitative Methods used for Data Collection and Analysis as well as Interpretive Triangulation with Previous Quantitative Analysis	39
Figure 3.2. CATS Item 5, Targeted Concept of Newton's 3rd Law .....	53
Figure 3.3. Initial Thematic Map, Showing Two Main Themes .....	58
Figure 3.4. Developing Thematic Map, Showing Three Main Themes .....	59

Figure	Page
Figure 3.5. Final Thematic Map, Showing Two Main Themes .....	59
Figure 5.1. Interrelationship of Item Design, Student Reasoning, and Item Difficulty..	109
Figure 5.2. Simplified Representation of the Conceptual Assessment Framework of Evidence-Centered Design (Pellegrino et al., 2014).....	110

## ABSTRACT

Denick, Dana L. Ph.D., Purdue University, May 2015. Difficulty as a Concept Inventory Design Consideration: An Exploratory Study of the Concept Assessment Tool for Statics (CATS). Major Professor: Ruth Streveler.

The ability for engineering students to apply mathematic, scientific and engineering knowledge to real-life problems depends greatly on developing deep conceptual knowledge that structures and relates the meaning of underlying principles. Concept inventories have emerged as a class of tests typically developed for use in higher education science and engineering courses. Concept Inventories (CIs) are multiple-choice tests that are designed to assess students' conceptual understanding within a specific content domain. For example, the CI explored within this study, the Concept Assessment Tool for Statics (CATS) is intended to measure students' understanding of the concepts underlying the domain of engineering statics. High quality, reliable CIs may be used for formative and summative assessment, and help address the need for measures of conceptual understanding. Evidence of test validity is often found through calculation of psychometric parameters. Prior research has applied multiple theoretical measurement models including classical test theory and item response theory to find psychometric evidence that characterize student performance on CATS. Common to these approaches is the calculation of item difficulty, a parameter that is used to distinguish which items are more difficult than others.

The purpose of this dissertation study is to provide context and description of what makes some CI items more difficult than others within the content area of statics, based on students' reasoning in response to CATS items. Specifically, the research question guiding this study is: how does student reasoning in response to CATS items explain variance in item difficulty across test items?

Think-aloud interviews were conducted in combination with a content analysis of selected CATS items. Thematic analysis was performed on interview transcripts and CATS development and evaluation documentation. Two themes emerged as possible explanations for why some CATS items are more difficult than others: (1) a Direction of Problem Solving theme describes the direction of reasoning required or used to respond to CATS items, and may also provide some description of students' reasoning in response to determinant and indeterminate multiple-choice problems; and (2) a Distractor Attractiveness theme describes problematic reasoning that is targeted and observed as argumentation for incorrect CATS responses. The findings from this study hold implications for the interpretation of CATS performance and the consideration of difficulty in concept inventory design. Specifically, findings from this study suggest that item difficulty may be associated with complexity, relating to theories of cognitive load. Complexity as it contributes to item difficulty is not solely dependent on the content of the concept inventory item, but also may be due to the item design and context of the test question.



## CHAPTER 1. INTRODUCTION

### 1.1 Background Information

Concept Inventories (CIs) are a class of tests that are typically developed for use in science and engineering courses. They are multiple choice tests designed to measure students' conceptual understanding within a specific domain. For example, the CI explored for the purposes of this study, the Concept Assessment Tool for Statics (CATS) is intended to measure students' understanding of the concepts underlying the domain of engineering statics, such as force and moment. CIs can be used for both formative and summative assessment purposes, in that they can be employed to measure conceptual understanding throughout the learning process as a feedback system for instructors as well as at the end of a course to measure learning gains (Richardson, 2004). In higher education science and engineering, they are commonly used to measure learning gains relative to pedagogical intervention or to diagnose conceptual understanding and misconceptions. As with any form of assessment, the usefulness of CIs is dependent on how well they test what they are intended to measure and inferences made using test data.

### 1.2 Problem Statement

In addition to the ability to apply design principles to engineering problems, engineers need to possess a deep conceptual understanding of scientific and engineering principles as a basis for reasoning and problem solving (Pellegrino, DiBello, & Brophy,

2014; Streveler, Litzinger, Miller, & Steif, 2008). However, research on learning in engineering education often finds that it is difficult for students to develop robust conceptual understanding, and students' misconceptions or problematic reasoning often persists throughout and beyond an undergraduate engineering curriculum (Streveler, Brown, Herman, & Montfort, 2014). Compounding this issue, engineering faculty have noted difficulty in assessing conceptual understanding (Pellegrino et al., 2014), as most assessment in engineering instruction is focused on measuring procedural knowledge through problem-based tests (Streveler et al., 2011). Overall there is a need for reliable assessment of learning in engineering education, specifically high quality tests of conceptual understanding (Olds, Moskal, & Miller, 2005; Pellegrino et al., 2014).

### 1.3 Statement of Purpose

The use of concept inventories has grown in engineering education due to the recognition of their success in spurring improved instructional practices in physics education (Richardson, 2004). High quality, reliable CIs may be used as formative and summative tests, and address the apparent need for measures of engineering students' conceptual understanding. This dissertation study intended to build upon evidence of CATS as a high quality CI (Hansen & Steif, 2006; Jorion, James, Schroeder, & DiBello, 2013; Santiago-Román, 2009; Steif & Dantzler, 2005) and to contribute to a body of knowledge on test development methods for high quality CIs as an example of a qualitative approach to evidence of validity.

The purpose of this dissertation study was to build upon an aspect of test design that makes meaning of test scores. In doing so, an exploratory qualitative study was used to elaborate upon implications and limitations of existing statistical models associated

with CATS. Prior research had determined psychometric measures that characterize student performance on CATS based on multiple theoretical measurement models, including classical test theory and item response theory. Common to these approaches is the determination of *item difficulty*, a parameter that is used to distinguish which items are more difficult than others. This dissertation study was conducted to provide rich description of what makes some CATS items more difficult than others within the context of statics and based on students' reasoning in response to CATS items.

#### 1.4 Research Question

The following research question guided this dissertation study:

- How does student reasoning in response to CATS items explain variance in item difficulty across test items?

In order to understand the underlying reasoning that accounts for CATS item difficulty, the use of qualitative methods allows for the collection of rich description and contextual detail that students may bring to their conceptual reasoning (Ericsson & Simon, 1993). It seems reasonable to expect a relationship between item difficulty and how students think about specific items. However, it is important to acknowledge a key assumption that it is not only possible to infer students' reasoning, but that the factors contributing to how students think about difficult assessment items are observable. This dissertation study employed qualitative methods to explore student reasoning in response to CATS items as a means of understanding why some CATS items are more difficult than others.

## 1.5 Overview of Methodology

CATS is a twenty-seven item multiple-choice concept inventory that was designed to diagnose students' correct and incorrect understanding of statics concepts (Steif & Dantzer, 2005). Previous studies evaluating CATS have collected large data sets of student responses and used statistical models such as classical test theory and item response theory to identify more difficult and less difficult CATS items (Jorion, James, Schroeder, & DiBello, 2013). This dissertation study built upon previous research by using qualitative methods of content analysis and think-aloud interviews to gather contextual information about the item design and student reasoning for each item. Eighteen undergraduate engineering students participated in the interviews and transcripts were created from audio recordings of the interviews. Using thematic analysis, verbal data were analyzed concurrently with a content analysis of CATS items and design literature. This analytic approach was selected to allow for an emergent thematic scheme that would address the research question and provide context and description as to why some CATS items are more difficult than others.

## 1.6 Rationale and Significance

In the development of learning assessment, it is important to apply systemic approaches and procedures that are grounded in educational research and practice regarding how people learn (Bransford, Brown, Cocking, Donovan, & Pellegrino, 2000). The Assessment Triangle is one such framework that can be applied in the development and evaluation of tests (Pellegrino, Chudowsky, & Glaser, 2001). The Assessment Triangle, as shown in Figure 1.1, models three interrelated elements of assessment: (1) Cognition, (2) Observation, and (3) Interpretation.

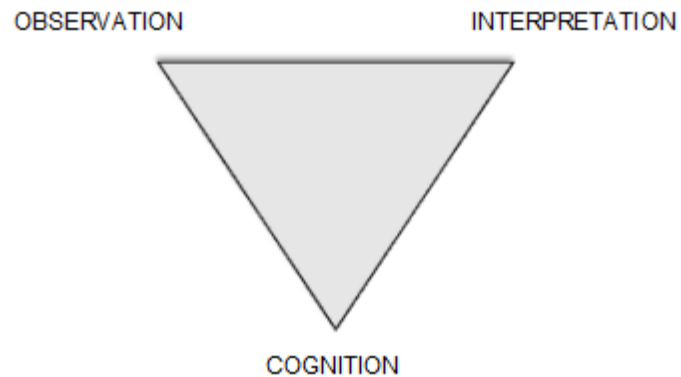


Figure 1.1 The Assessment Triangle (Pellegrino et al., 2001)

The cognition vertex of the assessment triangle describes how students demonstrate knowledge and develop competence within a domain (Pellegrino et al., 2001). For the present study, we can think of this element as a conceptual framework for statics that includes underlying concepts, common errors, skills, misconceptions, and other typical ways in which students learn statics and gain expertise in the subject area. The observation vertex includes the kinds of tasks that can prompt students to demonstrate desired knowledge (Pellegrino et al., 2001). In this case, the CATS items serve as the observation element. The interpretation vertex includes a model of what the evidence collected from the test means (Pellegrino et al., 2001). Since any test has some error in how it measures the actual knowledge that an individual holds about a topic, the interpretation element of the assessment triangle makes meaning from the test scores. For large scale test administration this often exists as a statistical model that characterizes student performance. However, qualitative interpretation models are also important as they provide contextual considerations that are not easily captured by a statistical model

(Douglas & Purzer, 2015). For this dissertation study, the interpretation element of the assessment triangle can be considered the interpretive argument for CATS that explains the meaning of student scores on CATS items and what decisions can be made based on those scores.

The primary focus of this study was to gather additional evidence to strengthen the interpretation vertex of the assessment triangle for CATS. However, due to the interrelated nature of the assessment elements, one may expect that exploring any one vertex will lead to implications for the other vertices. In particular, it is expected that this work will contribute to understanding not only what CATS items and statics concepts are difficult, but why they are difficult. This is a critical aspect for the cognition vertex of the assessment triangle, and an aspect that is often overlooked in the development of concept inventories (Streveler et al., 2011).

### 1.7 Organization of Dissertation

This dissertation is organized into five chapters. The first chapter provides a basic introduction to the study and briefly presents the motivation for the study, its purpose, research question, rationale and research approach. Chapter 2, the literature review, describes the research landscape in which this study is situated. The chapter offers background on the Assessment Triangle and its application to concept inventories in general and CATS in particular. Chapter 2 also includes a discussion on concept inventory development and a proposed gap in knowledge that this study addresses. Chapter 3 addresses the research design, data collection and analysis techniques used in this study. It describes the test design and evaluation literature examined as part of a content analysis of CATS items as well as the think-aloud interviews performed to collect

evidence of student reasoning in response to CATS items. Chapter 3 also includes a discussion on the use of thematic analysis and the process by which a thematic scheme was identified to describe the phenomena of interest. This chapter defines the themes and codes used to analyze the collected data. Chapter 4 presents the findings of the dissertation study through identification of CATS item coding according to the developed thematic scheme. This chapter also presents evidence from the collected data supporting the assignment of specific codes. Chapter 5, the discussion, concludes this dissertation. It discusses the findings and offers reasoning as to what the findings mean and their implications. Chapter 5 also includes the resulting theory of difficulty for CATS as a means of describing implications of this study for both instructors using CATS and concept inventory designers.

## CHAPTER 2. LITERATURE REVIEW

### 2.1 Introduction

There is growing interest in the improvement of teaching and learning assessment within higher education as evident by increasing occurrences of federal initiatives and heightened scrutiny of assessing outcomes of educational programs, particularly in science and engineering disciplines (Pellegrino et al., 2001; Pellegrino et al., 2014). Simultaneously, there is a growing trend within science and engineering education communities to develop, evaluate and validate tests designed to be used by instructors in classroom settings (Pellegrino et al., 2001; Singer, Nielsen, & Schweingruber, 2012).

#### 2.1.1 Development of Engineering Concept Inventories

The emergence of CIs within higher education science and engineering can be traced to a progression of research within science education that was concerned with identifying preconceived beliefs held by students prior to instruction in a subject area, and the development of tests used to identify naïve conceptions (Duit & Treagust, 2003; Treagust, 1988). However, it is the development and broad dissemination of the Force Concept Inventory (FCI) which served as a signifier of the potential of CIs to spur reform in higher education science and engineering communities (Richardson, 2004). The FCI was designed to assess students' conceptions of force within a Newtonian model of mechanics through deceptively simple multiple choice questions that do not require



extensive use of formulas or quantitative analysis (Hestenes, Wells, & Swackhamer, 1992). Physics instructors were initially surprised at how poorly students performed on the FCI despite often successfully demonstrating mastery of higher-level problem solving skills; Eric Mazur of Harvard University is a high profile convert who used surprising FCI results as a catalyst for the development of instructional strategies that were more effective with regard to students' conceptual learning (Mazur, 1997; Richardson, 2004). Large-scale reform of instruction within physics education communities is often attributed to the impact of a meta-analysis reporting significantly higher learning gains as measured by the FCI with instruction implementing active learning strategies over traditional lecture-based instruction (Hake, 1998). While both the veracity of Hake's study as well as the strength of the FCI as a diagnostic assessment tool have been called into question, the FCI remains a widely-used assessment instrument in physics education and has served as a reference for subsequent CI development.

Spurred by the acknowledged influence of the FCI, researchers within engineering education also began development of CIs in the early 2000's (Evans et al., 2003; Evans et al., 2002). For detailed descriptions of CIs in use or development within higher education science and engineering up to 2008, Reed-Rhoads & Imbrie (2008), and Libarkin (2008) prepared status reports of CIs in engineering education and science education respectively. Table 2.1 includes a sample of CIs currently in use in higher education science and engineering.

Table 2.1 Selection of Concept Inventories Developed for use in Engineering Education

Instrument	Domain	Key Reference
Statistics Concept Inventory (SCI)	Introductory statistics	(Allen, 2006)
Materials Concept Inventory (MCI)	Introductory materials engineering	(Krause, Decker, & Griffin, 2003)
Concept Assessment Tool for Statics (CATS)	Engineering statics	(Steif & Dantzler, 2005)
Dynamics Concept Inventory	Engineering dynamics	(Gray et al., 2005)
Thermodynamics Concept Inventory	Thermodynamics (engineering)	(Midkiff, Litzinger, & Evans, 2001)
Thermal and Transport Concept Inventory (TTCI)	Heat transfer, thermodynamics, fluid mechanics (engineering)	(Streveler et al., 2011)
Heat and Energy Concept Inventory	Temperature, heat and energy	(Prince, Vigeant, & Nottis, 2012)
Signals and Systems	Linear signals and systems	(Wage, Buck, Wright, & Welch, 2005)

In addition to the apparent differences of content domain, these concept inventories also vary in the extent and number of revision iterations as part of their development. CATS is unique among these CIs for having undergone evaluation based upon multiple, multi-institution administrations. While most CIs have been evaluated with basic psychometric analysis including correlational analysis, classical test theory and item response theory, few have been evaluated using structural models such as factor analysis or diagnostic models such as with the Fusion model evaluation of CATS. Details of these psychometric evaluations can be found in chapter section 2.4.2. Additionally, CATS is unique among these CIs with regard to the depth of development and evaluation history, with extensive documentation.

### 2.1.2 The Assessment Triangle

Both large-scale tests and those developed for use in the classroom should be created and evaluated with respect to the three aspects of the assessment triangle

(Pellegrino et al., 2001). The Assessment Triangle (shown in Figure 2.1) provides a framework for describing the interacting elements of developing, implementing and interpreting assessment of learning; also included in this figure is a translation of how the assessment triangle applies to CATS.

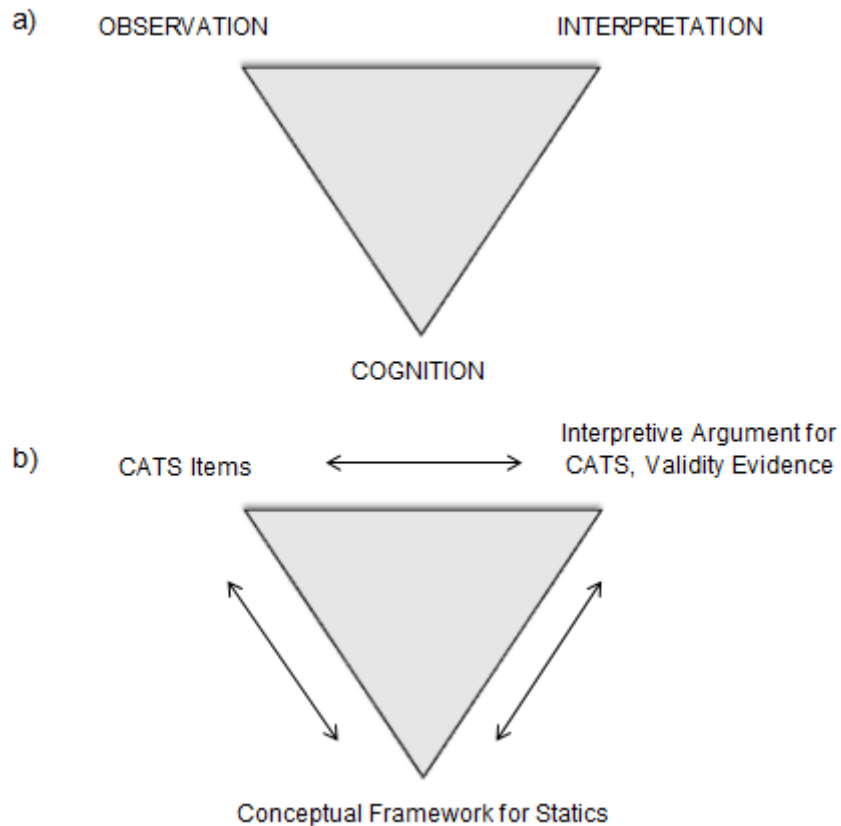


Figure 2.1 The Assessment Triangle (Pellegrino et al., 2001) including (a) elements of the assessment triangle and (b) as applied to CATS

The cognition vertex of the assessment triangle describes how students demonstrate knowledge and develop competence within a domain (Pellegrino et al., 2001). In the context of CATS, the cognition corner translates as a conceptual

framework for statics that includes underlying concepts, common errors, skills, misconceptions, and other common ways in which students learn statics and gain expertise in the subject area. The observation vertex includes the kinds of tasks that can prompt students to demonstrate the desired knowledge (Pellegrino et al., 2001). In this case, the CATS items including the test item problem stems, diagrams, and multiple choice response options serve as the observation element. The interpretation vertex includes a model of meaning that can be made from the collected test evidence (Pellegrino et al., 2001). Since any test has some error associated with the manner in which the test measures the actual knowledge that an individual holds about a topic, the interpretation element of the assessment triangle makes meaning from the test data. For large data sets the interpretation vertex or inferences made from test scores often takes form as a statistical model that characterizes student performance. However, qualitative interpretation models are also important as they provide contextual considerations that are not easily captured by a statistical model. In the context of CATS, the interpretation element of the assessment triangle can be considered the interpretive argument for CATS that explains the meaning of student scores on CATS items and what decisions can be made based on those scores. This element also includes validity evidence that supports the adequacy and appropriateness of the interpretive argument. Although validity evidence is often described as applying to the test itself, this is incorrect. Validity is better described as the extent to which empirical evidence and theoretical rationale support the inferences made on assessment results (Messick, 1990). “To validate a test-score interpretation is to support the plausibility of the corresponding interpretive

argument with appropriate evidence (Kane, 1992, p. 527).” A more extensive description of assessment validity can be found in chapter section 2.4.3.

While there has been considerable effort within science and engineering education communities to characterize key concepts and misconceptions within specific domains – the cognition vertex of the assessment triangle, and develop concept inventories for use in the classroom – the observation vertex of the assessment triangle, less effort has been focused on the interpretation vertex of the assessment triangle (Pellegrino, DiBello, Jorion, James, & Schroeder, 2013). Specifically, there is need for in-depth examination of what students know and how they know it contained within patterns of test scores (Pellegrino et al., 2013). The following sections within this review of literature provide an in-depth discussion of how each vertex of the assessment triangle can be applied to an enhanced understanding of CIs for engineering education and CATS in particular.

## 2.2 Expanding upon the Cognitive Element of the Assessment Triangle

As previously stated, the cognitive vertex of the assessment triangle includes all aspects of how learning occurs within the domain of interest. “A central premise is that the cognitive theory should represent the most scientifically credible understanding of typical ways in which learners represent knowledge and develop expertise in a domain (Pellegrino et al., 2014).”

### 2.2.1 Nature of Conceptual Understanding

Concept Inventories are multiple choice tests designed to measure conceptual knowledge within a specific domain. A useful definition of conceptual knowledge is the “implicit or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain (Rittle-Johnson et al., 2001, p.

346).” Additionally, conceptual knowledge can be described as abstract or generalizable (Rittle-Johnson et al., 2001). From a discrete perspective, “concepts are pieces or clusters of knowledge, for example, force, mass, causation, and acceleration (Streveler et al., 2014). The perspective I have taken for describing conceptual knowledge is a networked understanding of interrelated concepts, which is subject to a constructivist approach to learning wherein new knowledge is incorporated into existing networks of prior knowledge (Hiebert & Lefevre, 1986).

I would be remiss in discussing conceptual knowledge and understanding without acknowledging the discrepancy in views of concept structure. There are schools of thought that concepts are conceived as theory-like structures describing a comprehensive worldview, while other perspectives view concepts as individual pieces of knowledge that are reorganized based on experience. The pieces versus coherence argument in conceptual literature may seem incommensurable; however I contend that cognitive models of how knowledge is structured do not necessarily need to be mutually exclusive. The wealth of research on conceptual understanding in STEM fields prohibits a comprehensive review of literature; however I have selected to cover pertinent resources that can inform research on conceptual understanding of statics.

Within the theory-like strand of conceptual research, similarities were drawn between students’ conceptualization of force and historical impetus theories (Clement, 1982; McCloskey, 1983). The main tenets of impetus theory can be described as the tendency to view motion in general as evidence of the presence of a force, holding the idea that an object is only in motion if there is a greater force in that direction than in any opposing direction, and that forces will gradually increase or decrease depending on the

motion of an object and its interaction with other objects or forces such as gravity (Clement, 1982; McCloskey, 1983). This perspective is valuable as impetus theory can be used to describe common misconceptions that students hold regarding force, and offer some way of generalizing these common misconceptions. However, as these authors note, there are differences in student conceptions that cannot be captured by impetus theory alone.

The knowledge-in-pieces strand of conceptual research is characterized by breaking down the conceptual understanding of a domain into bits of knowledge that may be interrelated to one another. An example of this structure is diSessa's (1993) itemizing of students' phenomenological primitives (p-prims) described as pieces of knowledge that students exhibit and that result from interaction with their external environment. These pieces of knowledge such as Ohm's p-prim, which describes how forces are needed to overcome a resistance, are themselves correct, but are not always applied to context correctly (diSessa, 1993). From this perspective, the misapplication of p-prims is at the heart of misconception (Hammer, 1996). Similar research has sought to identify facets of knowledge that are not at the indivisible granularity of p-prims, but are clumps of knowledge that characterize student thinking, such as "horizontal movement makes a falling object fall more slowly" (Minstrell, 2000). Viewing conceptual understanding as knowledge-in-pieces can be useful as the foundation for assessments that measure key concepts and misconceptions, while also allowing for individual differences in how concepts are understood (diSessa, 2008).

In some ways external to the pieces versus coherence argument, there are additional theories of conceptual understanding that have informed this dissertation study.

In response to research on naïve physics, Vosniadou (2002) proposed that individuals' observations and beliefs about phenomena as informed by everyday experience are also related to underlying epistemological and ontological factors that in combination form a framework of conceptual understanding. This perspective is useful in that misconceptions may not always be corrected due to a change in the concept itself, but may also require changes to beliefs regarding the nature of knowledge and the process of knowing (Stathopoulou & Vosniadou, 2007). Some researchers have also noted a trend in student misconceptions due to an incorrect categorical organization of knowledge; for example force may be misconceived as a substance and lead to incorrect thinking that relates force to material (Reiner, Slotta, Chi, & Resnick, 2000). Additionally, students may incorrectly assume that emergent processes are the result of a sequential or 'cause and effect' causal structure, instead of as the result of random interactions (Chi, Roscoe, Slotta, Roy, & Chase, 2012). Using the language of Chi et al. (2012), sequential processes are those that have a causal relationship such as a force being applied to an object that causes an acceleration in the direction of motion. Emergent processes are those that are acausal (non-linear and cannot be attributed to the action of any one agent), and may be described as macro-level patterns of lower-level interactions (Chi et al, 2012). For example, static equilibrium may be considered an emergent process as a system in static equilibrium is characterized by forces and moments in balance at all points of the system simultaneously and independently.

### 2.2.2 Statics Concepts and Conceptual Framework

Statics courses often serve as an introduction to how engineers model real-world systems, and contain a marked increase in complexity compared to introductory physics



(McCracken & Newstetter, 2001). The increased complexity of statics can be accounted for by the need to consider both force and moment balances (Litzinger et al., 2010; Newcomer & Steif, 2008a; Ortiz, Heron, & Shaffer, 2005), as well as the need to examine connections between and among multiple bodies within mechanisms (Steif & Dantzer, 2005). Despite the difference in complexity between introductory physics and statics, findings from research across domains can be drawn upon for identification of students' conceptual understanding and problematic reasoning. Findings have suggested that students have difficulty recognizing that forces act in pairs between bodies, and do not exist as individual entities (Brookes & Etkina, 2009; Minstrell, 2000; Steif & Dantzer, 2005). Students have difficulty recognizing when concepts of moment are required to understand the static conditions of a system (Newcomer & Steif, 2008a; Ortiz et al., 2005). Additionally, students have difficulty understanding how forces, couples, and moments may interact to produce a static condition, or balance (Newcomer & Steif, 2008c).

In an effort to articulate an underlying static concepts, Steif (2004a) drew upon his experience as a statics instructor to put forward the argument that a conceptual framework underlying the domain of statics encompassed more than the equilibrium principle. By describing the nuances of physical reasoning that students must consider when implementing the equilibrium principle in statics problems, Steif (2004a) developed an initial conceptual basis for statics that included: (1) the nature of forces acting between bodies, (2) using single force or couple to represent distributed forces, (3) the nature of the contact implies specific types of simplified forces, and (4) the zero net conditions of force and moment for equilibrium conditions. Although not used a guiding

reference in CATS development at the time, this initial approach to understanding the conceptual framework of statics can be viewed as the initial attempt to explain the cognition element of the assessment triangle. A revised version of the conceptual framework for statics is shown in Table 2.2

Table 2.2 Clusters of Concepts for the Conceptual Framework of Statics (from Steif & Dantzler, 2005, p. 363)

Cluster	Description
1	Forces are always equal and opposite pairs acting between bodies, which are usually in contact.
2	Distinction must be drawn between a force, a moment due to a force about a point, and a couple. Two combinations of forces and couples are statically equivalent to one another if they have the same net force and moment.
3	The possibilities of forces between bodies that are connected to, or contact, one another can be reduced by virtue of the bodies themselves, the geometry of the connection and/or assumptions on friction.
4	Equilibrium conditions always pertain to the external force acting directly on a chosen body, and a body is in equilibrium if the summation of forces on it is zero and the summation of moments on it is zero.

In addition to the conceptual framework put forward through the identification of concept clusters, Steif drew from his experience as a statics instructor and through iterations of item responses to create a list of conceptual errors that students make in statics (Table 2.3).

Table 2.3 Conceptual Errors in Statics (from Steif &amp; Dantzler, 2005, p. 364)

Error	Description
1	Failure to be clear as to which body is being considered for equilibrium.
2	Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two.
3	Leaving a force off the free body diagram (FBD) when it should be acting.
4	Drawing a force as acting on the body in the FBD, even though that force is exerted by a part which is also included in the FBD.
5	Drawing a force as acting on the body of the FBD, even though that force does not act directly on the body.
6	Failing to account for the mutual (equal and opposite) nature of forces between connected bodies that are separated for analysis.
7	Ignoring a couple that could act between two bodies or falsely presuming its presence.
8	Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces.
9	Presuming a friction force is at the slipping limit ( $\mu N$ ), even though equilibrium is maintained with a friction force of lesser magnitude.
10	Failure to impose balance of forces in all directions and moments about all axes.
11	Having a couple contribute to a force summation or improperly accounting for a couple in a moment summation.

Although there has been substantial attention paid to identifying examples of errors that students make in statics and in response to CATS items, there is a need for additional research that not only inventories the errors students make, but explains *why* they make them.

### 2.3 Expanding upon the Observation Element of the Assessment Triangle

The observation element of the assessment triangle represents the specific tasks that are used to collect evidence within assessment. Tests and test items should be carefully designed to provide evidence that can link cognitive elements grounding the

assessment instrument with the interpretation of test scores (Pellegrino et al., 2014). This applies to intentional design of the entire as well as to the individual test items.

### 2.3.1 Concept Inventory Item Design

Concept Inventories are typically composed of multiple choice items that are intended to measure students' ability to apply a concept or specific piece of knowledge in an accurate manner. Due to differences across domains, there is no uniform idea of the size and scope of the target reasoning for concept inventory items (Lindell, Peak, & Foster, 2007). Some general rules of thumb include that CI items should measure a discrete concept or piece of knowledge, and incorrect responses should align with common errors or misconceptions within the domain (Streveler et al., 2011). Additionally, because CIs are intended to measure conceptual knowledge, it is common practice to include problems that do not require extensive mathematical or procedural reasoning in order to arrive at a correct solution (Richardson, 2004).

Figure 2.2 provides an example of the format typically used in concept inventories, including the key components of CI item design: a problem stem with any number of possible responses among which one is correct while the other incorrect responses are referred to as distractors.

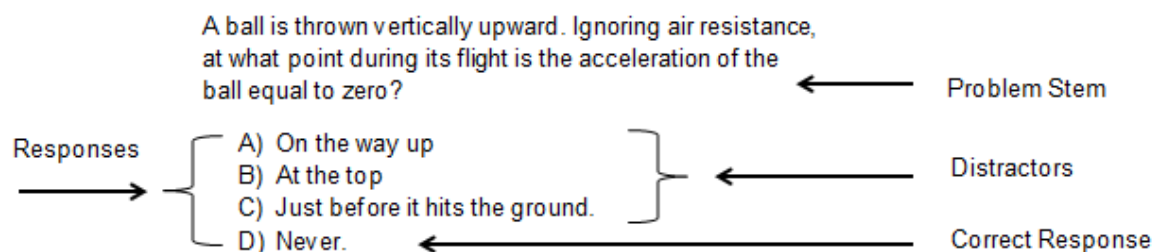


Figure 2.2 Concept Inventory Item Components

### 2.3.2 CATS Item Design

CATS employs a combination of worded problem statements and questions with extensive use of diagrams to convey pertinent information. This may be considered both a benefit and a detriment since the problems do not require high levels of language-based reasoning, but do involve significant demands of diagrammatic reasoning. An example CATS item is shown in Figure 2.3.

15. The mechanism is acted upon by the horizontal force shown. The pin in the slot is frictionless.

What is the direction of the force exerted by the slot on the pin A?

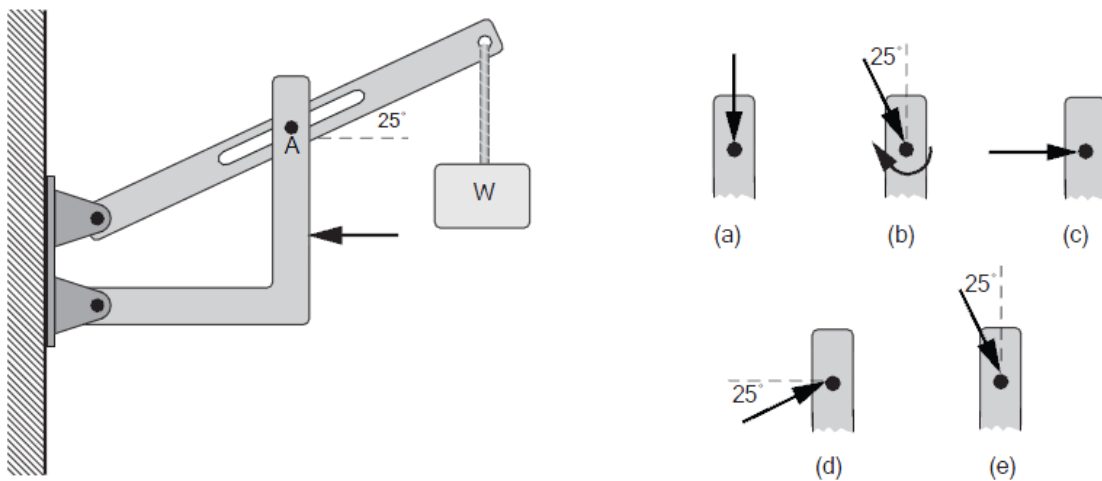


Figure 2.3 Sample CATS Item

As with other CI items, CATS was designed with the intention that each item target a specific and unique statics concept (Steif & Dantzler, 2005). The concept addressed by each item comprising CATS is presented in Table 2.4.

Table 2.4 Concepts addressed on each item (from Steif &amp; Hansen, 2007, p. 206)

Concept	Name	Description	Items
A	Drawing forces on separate bodies	Identifying forces acting on a subset of a system of bodies.	1 – 3
B	Newton's 3 <sup>rd</sup> Law	Forces between two contacting bodies must be equal and opposite.	4 – 6
C	Static Equivalence	Static equivalence between forces, couples, and combinations.	7 – 9
D	Roller joint	Direction of force between the roller and the rolled surface.	10 – 12
E	Pin-in-slot joint	Direction of force between pin and slot of a member.	13 – 15
F	Loads at surfaces with negligible friction	Direction of force between frictionless bodies in point contact.	16 – 18
G	Representing loads at connections	Representing unknown loads at various connections.	19 – 21
H	Limits on friction force	Sorting out implications of equilibrium and Coulomb's Law of friction force.	22 – 24
I	Equilibrium	Consideration of both force and moment balance in equilibrium.	25 – 27

### 2.3.3 Concept Inventory Development Process

Typically, concept inventories are developed in a manner that aligns with some aspects of recommendations for assessment design from educational research. Primarily, assessment design should consider its purpose, whether to assist in learning, measure individual ability or attainment, or to evaluate a program (Pellegrino et al., 2014).

Additional considerations include the context in which the assessment is administered and any practical constraints such as resources and time available to the test designer.

The process by which concept inventories are developed does not vary greatly from recommendations for multiple choice test development (Haladyna, 2012).

Haladyna (2012) recommends that items, response formats, scoring procedures and test administration procedures should align with test purpose, domain and intended test-taking populations. Concept inventories are primarily associated with higher education science and engineering and as previously noted, these should be grounded in cognitive theories of how students learn in these domains and should also be appropriate for the intended test use. Also, just as with any multiple choice test, the process used to develop concept inventories should be documented. This includes any frameworks used to develop and revise items, the decision-making process for item selection, and the psychometric evaluation of the test and test items including description of sample population, characterizing parameters, and appropriateness of model (Haladyna, 2012). Haladyna (2012) also emphasizes the relationship between the intended test-taking population and the test development. For example, pilot populations should be as representative as possible of the intended test-taking population and should be documented, and test developers should seek to detect and eliminate any aspect of test design that might bias test scores for particular populations.

#### 2.3.4 Development of CATS

As part of an effort to better understand fundamental concepts underlying statics as well as common errors and misconceptions of students, in the early 2000's Paul Steif and collaborating researchers began development of a statics concept inventory (Steif, 2003). As previously noted, this endeavor was concurrent with an overall trend in engineering education to develop concept inventories as test that would empower instructors to measure students' learning (Evans et al., 2003). Just as the FCI spurred educational reform in physics, engineering education researchers sought to develop tools

that would assess students' conceptual understanding as an indicator of learning for the main purpose of comparing instructional approaches (Evans et al., 2002).

An initial version of the CI was administered during the 2003-2004 academic calendar; analysis from this initial administration of items in development for the statics concept inventory was presented as a means of exploring item performance (Steif, 2004b). At this stage, Steif worked with colleagues who were also experienced statics instructors to build out the cognitive framework of statics to include skills that students would need mastery of in order to correctly apply statics concepts, and common errors that students make with statics problems (Steif, 2004a, 2004b). In addition to the expert's point of view, errors were identified through student interviews with items developed for early versions of the instrument (Steif, 2004b). Items that did not perform well as measured by item analysis were dropped, new items were developed for inclusion into the CI and some existing items were modified based on psychometric evaluation of the CI. This process was repeated in the 2004-2005 academic year; the version of the CI prepared for 2005-2006 administration was expected to only require minor modifications (Hansen & Steif, 2006)

As an illustration of the process used in the development and evaluation of CATS, Santiago-Roman (2009) constructed the flowchart shown in Figure 2.4 with indication of the feedback loops used to inform the iterative development of engineering CIs.



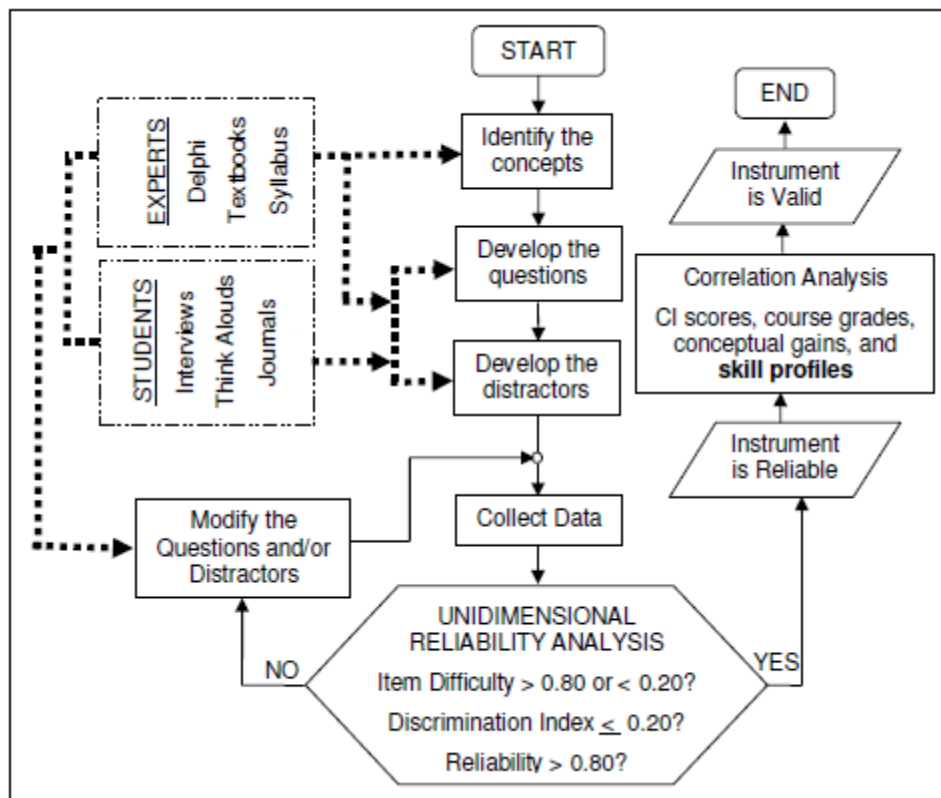


Figure 2.4 Problematic Concept Inventory Development Process Flowchart, (Santiago-Román, 2009)

This flowchart shows the iterative nature and variety of sources of evidence used in the development and revision of CATS items. However, it also contains some key limitations. Specifically, this model includes purely quantitative evidence of validity and the qualitative evidence of student reasoning does not feed back into the conceptual framework of the development process. Most glaringly, this process model implies that there is an “end” to the development of CI design. As pointed out in previous research, the development of CIs is a never-ending process (Streveler et al., 2011). Continuing improvement efforts allow for the development of additional tests and items with increased alignment to assessment goals. Additionally, varying populations may require

adaptation of CIs for specific population needs, whether cognitively or culturally-based.

Figure 2.5 shows a modification of this CI development flowchart addressing the key limitations of solely quantitative evidence of validity and implying that there is an end to test development.

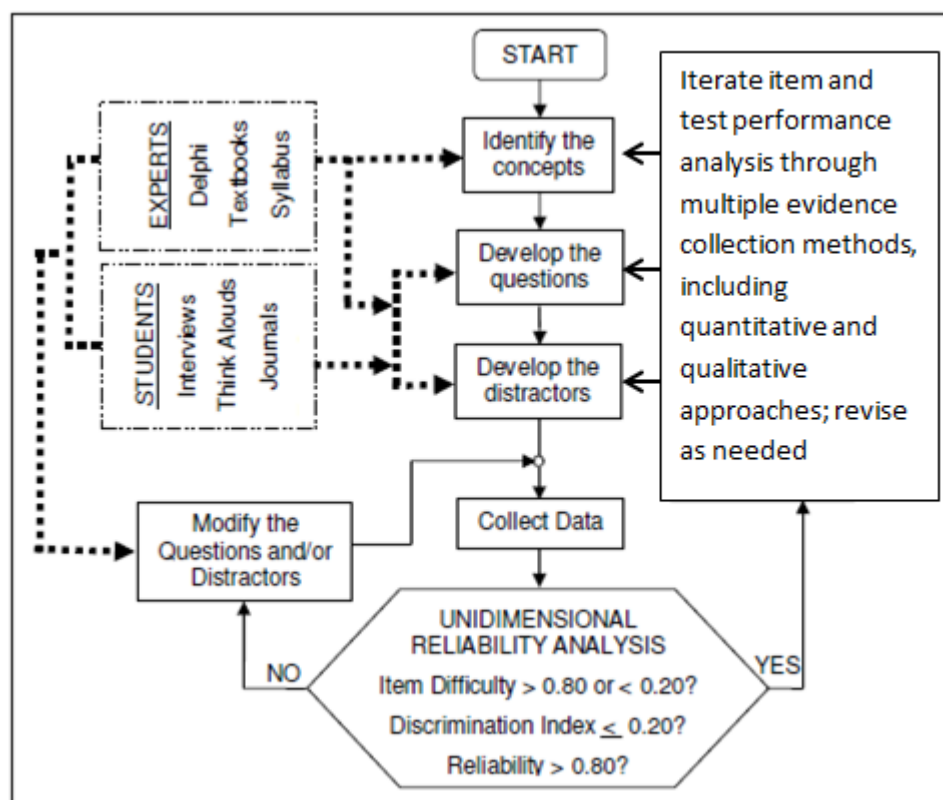


Figure 2.5 Revised Concept Inventory Development Process Flowchart, modified from

(Santiago-Román, 2009)

#### 2.4 Expanding upon the Interpretation Element of the Assessment Triangle

The interpretation vertex of the assessment triangle includes all of the methods and tools used to reason from fallible observations. This includes a way of explaining the meaning behind test scores and decisions that can be made based on the collected evidence. Typically, the interpretation vertex of the assessment triangle is addressed by

collecting large sets of test responses and using statistical methods to produce psychometric models that describe students' test performance. Additionally, common approaches to interpretation of test scores include quantitative measures of correlation to other known performance indicators, measure of validity such as goodness of fit for factor analysis, and reliability measures such as Cronbach's alpha. Less frequently, qualitative approaches are used to provide descriptive evidence of what test scores mean.

#### 2.4.1 Psychometric Evaluation

Test theories allow researchers to apply a model or framework that links observed measures to latent characteristics (Hambleton & Jones, 1993), such as how to relate a test score to some meaning of what is represented by that test score. As previously described, the test theories that undergird psychometric analysis align with the interpretation element of the assessment triangle in that they provide meaning to test scores. Classical Test Theory (CTT) and Item Response Theory (IRT) are the main test theories that are applied as part of psychometric evaluation. Both of these approaches are statistical theories that explain variance in patterns of test scores (Haladyna, 2012). As with all models, CTT and IRT provide incomplete representations of test measures and include some amount of error. However, models with good fit to test score data can be used to feedback into item design to produce tests with desired parameters (Hambleton & Jones, 1993).

Classical Test Theory primarily models measures at the test-level, however it does include two important item-level parameters: *item difficulty* and *item discrimination*.

- *Item difficulty* expresses the proportion of students who answer an item correctly.

Item difficulty can range from 0 to 1, with 0 indicating no correct responses and 1

indicating all correct responses. A range of item difficulty is desired for an assessment as a whole, but recommendations suggest that individual item difficulties should range between .20 at the most difficult end and .80 at the least difficult end (Haladyna, 2012).

- *Item discrimination* expresses how well the item serves to discriminate between higher and lower levels of ability. For example if an item is answered correctly by most of the higher level group and few of the lower level group, it would have a high item discrimination. Item discrimination also ranges between 0 and 1, and a general rule is that item discrimination should be above .20 (Haladyna, 2012).

Due to the nature of CTT models, this statistical approach is sample-dependent. CTT is also dependent on key assumptions: the test score error is uncorrelated with the actual test score, the average test score error across the population sample is zero, test score error for tests that measure the same content and would produce the same score is also uncorrelated (Hambleton & Jones, 1993).

Item Response Theory assumes that variance in test response patterns can be modeled at the item-level. Various statistical models may be applied that link underlying abilities with specific item scores. This often takes the form of one-, two-, or three-parameter logistic functions that link observed measurement with latent characteristics performance, or test scores with assessment meaning (Hambleton & Jones, 1993). A key added value of IRT in relation to CTT is that individual items are linked to specific underlying abilities. In the case of CATS, this would mean that IRT analysis would link individual items with specific statics concepts. Additionally, IRT is a predictive measure

in that the *test characteristic function* allows for a prediction of test scores for a given ability level. Three-parameter IRT models include consideration of function variables that correspond with *item difficulty*, *item discrimination* and a pseudoguessing variable that accounts for performance of low-ability test takers on multiple choice assessments. A two-parameter IRT model neglects the pseudoguessing variable, while the one-parameter IRT model sets *item discrimination* as a constant. IRT requires a larger sample size to calibrate the statistical model and determine goodness of fit statistics, but unlike CTT is independent of the sample population.

#### 2.4.2 Psychometric Evaluation of CATS

Early development of CATS involved studies that applied basic psychometric techniques including correlating performance on items in development for CATS with traditional problem solving (Steif, 2003), test and item analysis identifying values of item difficulty and item discrimination (Steif & Dantzler, 2005), and factor analysis (Hansen & Steif, 2006). Subsequent research has determined psychometric measures that characterize student performance on CATS based on multiple theoretical measurement models including test and item performance analysis through CTT and IRT, structural analysis through factor analysis, subscale reliability and tetrachoric correlations (Jorion, James, Schroeder, & DiBello, 2013), and diagnostic modeling through the Fusion Model fit of a Q-matrix (Santiago-Román, 2009).

#### 2.4.3 The Interpretation Vertex as a Validity Argument

Current views on assessment validity include the notion that validity depends on the extent of supportive evidence for specific interpretations of test scores (Haladyna, 2012; Kane, 1992; Messick, 1990). Validity is not a measurement or absolute value; “the

best that can be done is to show that the interpretive argument is highly plausible, given all available evidence (Kane, 1992, p. 527).” Traditionally, a common approach to demonstrating evidence of validity is to attend to forms of validity that in aggregate provide a larger validity argument: content validity provides evidence that the content of the test aligns with the desired domain or subject matter; criterion validity provides evidence that the test scores align with other comparable measures; construct validity provides evidence that the test measures what it intends to measure with regard to concept, skill, or construct; consequential validity provides evidence of the use of the assessment tool and its impacts on instruction and learning (Messick, 1990, 1994).

#### 2.4.4 A Comprehensive Approach to Validity for Concept Inventories

As part of on-going research to develop a comprehensive approach to validity for CIs in engineering education, Pellegrino and colleagues (2013) describe parameters of a validity model that can be considered in the context of classroom assessment. These include *cognitive aspects of validity*, *instructional aspects of validity*, and *statistical aspects of validity*. Pellegrino (Pellegrino et al., 2013) contends that these aspects of validity are inclusive of and more specific than the traditional forms of content, criterion, construct and consequence validity. Validity from a cognitive perspective considers to what extent an instrument targets the cognition or understanding of desired knowledge, distinct from confounding cognitive factors such as language (Pellegrino et al., 2013). For CIs as tests of conceptual understanding, this can be interpreted as making sure that the CI is tapping into the desired forms of conceptual knowledge and misconceptions. Instructional aspects of validity relate to the use of CIs for formative assessment and consider to what extent a test provides useful and relevant information for instructional

integration (Pellegrino et al., 2013). This requires evaluating the alignment of course content and the conceptual domain of a specific CI as well as the relevance of the CI with regard to specific curricular goals. The statistical aspects of validity include how well the inferences made about student understanding are empirically supported, and “to what extent does an assessment reliably yield model-based information about student performance, especially for diagnostic purposes (Pellegrino et al., 2013, p. 8).” As previously mentioned, the usefulness of psychometric models for learning measures depends on the appropriateness of the linkages between test scores and the underlying abilities intended to be measured by the test (Hambleton & Jones, 1993). In the context of CIs, this relates to the strength of psychometric and statistical models to provide information regarding students’ conceptual understanding, specifically as a means of informing instructional practice to improve student learning.

The means by which evidence of validity is collected within the comprehensive approach to validity for CIs in engineering education includes multiple, related modes of data collection and analysis that examine evidence of validity for specific tests, scoring procedures, and support of test score inferences (Pellegrino et al., 2013). Figure-2.6 shows the interactive nature of data sources and collection activities. Table 2.5 describes how the interrelated data collection strategies align with aspects of a validity argument.

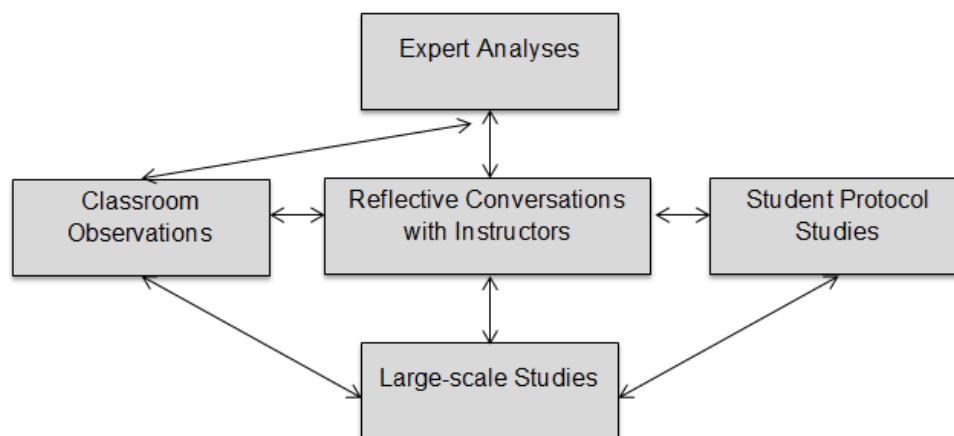


Figure 2.6 Proposed Data Sources and Collection Strategies as part of a Comprehensive Approach to Validity (adapted from (Pellegrino et al., 2013))

Table 2.5 Relation of Data Collection to Validity Components as part of a Comprehensive Approach to Validity (adapted from (Pellegrino et al., 2013))

Data Collection Activity	Cognitive Validity	Instructional Validity	Psychometric Validity
Expert Analyses & Discussions with Instructors	Is design supported cognitively with respect to critical forms of knowledge and understanding?	Does design support instructional needs and uses and instructor understanding?	Is design supported psychometrically with respect to adequacy of scoring and measurement?
Cognitive Protocol Studies	<b>What students actually do while working on assessment activities; what is being probed by the activity; what scoring &amp; feedback reveal.</b> Issues regarding linguistic and cultural diversity.	Degree to which test and item scores are expected to interact reasonably with other instructional indicators and benchmarks.	<b>Provide confirmation for interpreting model parameters, covariance analyses, and relationships to other variables.</b>
Discussions with Instructors & Classroom Observations	Instructor perceptions of cognitive information provided by assessment activities – what is revealed about their students' understanding; what responses would be expected. Student opportunities to learn as related to assessment activities.	Instructors' understanding and use of embedded assessments overall, to guide and differentiate instruction. Fidelity of assessment use.	Instructor knowledge about the use of score reliabilities, item difficulties, expected student responses and variations. Variability in conditions of assessment administration.
Large Scale Studies of assessments and external test performance	Variability in item responses, and aggregated scores, test and item information functions. All these relative to cognitive aspects of student performances.	Relationship to instructional needs and utility of measures such as reliabilities, parameter estimates, item analysis, variability. Monitor class performance progression.	Reliability of test scores, and extended to diagnostic tests. Model-data fit. Factor & dimensionality analyses. Differential functioning for student linguistic and ethnic groups. Predictive validity; alignment with other measures.



Of particular interest to this study is how the role of protocol or think-aloud studies are incorporated into the comprehensive approach to validity. While this approach clearly illustrates the role of think-aloud studies as a source of qualitative evidence that can feedback into all of the assessment triangle vertices, there is a potential opportunity not included in this approach wherein think-aloud studies not only confirm the interpretation of model parameters, but also provide detailed context for what they mean.

## 2.5 Addressing Gaps in Knowledge

The current state of understanding item difficulty as it relates to concept inventories relies on quantitative measures, with item difficulty values found through application of psychometric models. These statistical models are effective at identifying difficult items based on how students respond to test items, but are not necessarily based on why students select specific responses. Theories of conceptual understanding provide some basis for how students may respond to items from a conceptual perspective, but do not provide sufficient explanation as to why some conceptual test items are more difficult than others. In order to understand the relationship between CI item design and students' reasoning in response to items, additional literature may address some of these limitations.

### 2.5.1 Cognitive Load Theory

With regard to rationale that explains the difference of performance on CI test items, cognitive load theory may have potential to connect student's knowledge or ability to their test performance. Cognitive load theory describes the finite cognitive capacities available to an individual as cognitive loads are imposed on working memory (Sweller et al., 2011). This idea relates to educational assessment in that assessment tasks that

require too much cognitive capacity may hinder students' ability to demonstrate intended performance, ability, or knowledge (de Jong, 2010).

Research on cognitive load as it applies to learners find that novice learners have incomplete knowledge structures that require additional loads on working memory to comprehend concepts in a domain and apply them in context; while experts have routinized much of their thinking and do not tax their working memory when posed with the same problems (Sweller, 1988). This may explain why CI items that are seemingly trivial to domain experts prove difficult for undergraduate students.

### 2.5.2 Evidence-Centered Design

The assessment triangle as a guiding framework for test development and evaluation is useful in its rationale: that there needs to be alignment between cognition, observation, and interpretation. However, it can be difficult to understand how to apply the assessment triangle in context. Evidence-Centered Design (ECD) is a structured approach to test design that identifies layers of tasks and structures that contribute to the collection of evidence of students' performance. ECD assumes that there is inherent error in assessment, and assessment as a construct is similar to validity in that it is an argument based on incomplete or imperfect evidence (Mislevy & Haertel, 2006).

One aspect of ECD especially useful for applying the assessment triangle in context is the Conceptual Assessment Framework (Figure 2.7).

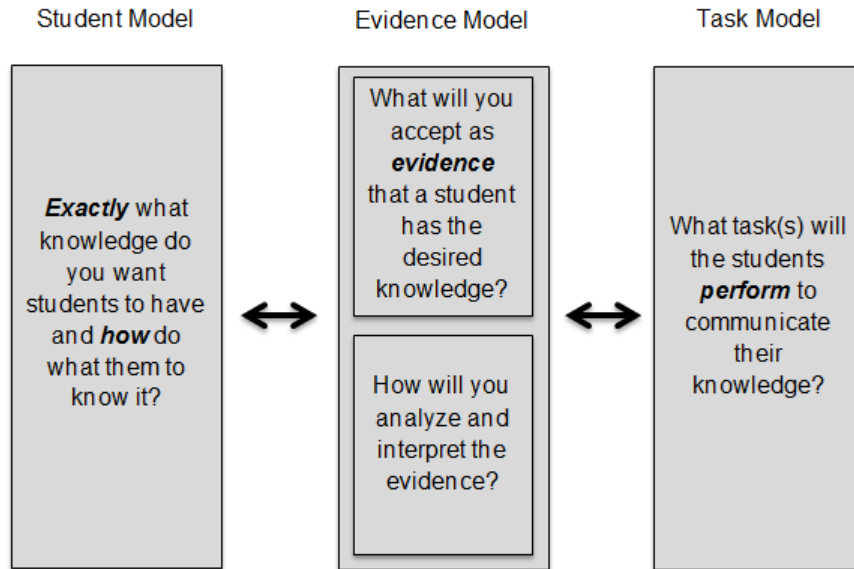


Figure 2.7. Simplified Representation of the Conceptual Assessment Framework of Evidence-Centered Design (Pellegrino et al., 2014)

The Conceptual Assessment Framework provides a basis from which assessment design decisions can be made:

- Student model describes what the assessment is intended to measure
- Task model describes the context in which evidence is collected through some form of student performance
- Evidence model describes rules and models for drawing meaning from student performance evidence and observations (Mislevy & Haertel, 2006; Pellegrino et al., 2014)

Another practical aspect of ECD is the adoption of a design pattern. A design pattern is a template that may be used by test developers to guide decision-making in test development by identifying key elements of an assessment argument. The design pattern

may include examples of test features such as rationale, focal knowledge, additional knowledge, potential work products, potential observations, characteristic features of tasks, variable features of tasks (Mislevy & Haertel, 2006). In the case of a concept inventory, a design pattern would include a structure conceptual framework, common characteristics among CI items, templates for problem stems aligning with specific concepts, examples of student reasoning in response to specific distractors that provide evidence of desired reasoning, and identification of additional item features that may contribute to comprehension, cognitive load, item difficulty, among other test parameters.

## 2.6 Summary

In summary, through a review of literature describing how the assessment triangle applies to concept inventory development with specific consideration of CATS, some key points have been identified. Specifically, the extensive documentation of the test development process and multiple evaluations of CATS provides a unique collection of evidence from multiple sources on the performance of CATS that may inform this study. Additionally, there are key gaps in knowledge regarding CATS and CI development that this study may help to address. Namely, there is a need for research that not only itemizes student errors, but provides explanation of students' problematic reasoning that leads to errors. Also, there is opportunity for qualitative evidence to potentially feedback into the conceptual framework of CI development and to provide meaning for quantitative psychometric measures within the context of statics.

### 2.6.1 Interconnectedness of the Assessment Triangle

Paramount to how each vertex of the assessment triangle applies to the context of CATS is the need for alignment among the assessment elements. This alignment can be

strengthened through iteration of evaluation as a means of enhancing an interpretive argument for validity. By conducting a qualitative analysis of CATS and building upon an interpretive argument, findings from this study will inevitably feedback into both the cognition and observation elements. Due to the interrelated nature of the elements of the assessment triangle, it is expected that exploring any one vertex will lead to implications for the other vertices. Although the present study used the interpretation element as a starting point, it was expected that findings will hold implications for CATS item design and the conceptual framework for Statics.

## CHAPTER 3. METHOD

### 3.1 Introduction

This chapter details the research design and methods used for the present study. In doing so, a review and rationale for the selection of research design, data sources, data collection approaches, and method of analysis are provided for the exploratory qualitative study that conducted. The purpose of this dissertation study is to provide a contextual explanation for the sources of difficulty in CATS items.

Specifically, the research question guiding this study is:

- How does student reasoning in response to CATS items explain variance in item difficulty across test items?

### 3.2 Research Design

Although the purpose of the present study was to explain or account for item difficulty variance, the nature of the study conducted is more exploratory. Since there are no existing theories of difficulty for concept inventories, an exploratory qualitative study was conducted as a means of developing such a theory, at least for CATS. Additionally, a qualitative approach was appropriate for this study because the research question requires the consideration of multiple forms of evidence that can be holistically evaluated (Creswell, 2008). The general approach taken with this dissertation study may be described as data-driven, in that data collected through a content analysis and think-aloud

interviews served as the source for the analytical system used to understand the data. Through thematic analysis, a coding scheme was developed that was then systematically applied to CATS items and student interview verbal data. The coded data was then triangulated with previous psychometric measures as a means of answering the guiding research question.

A summary of the research design is represented in Figure 3.1, including representation of the data sources within the data collection phase along with analytic methods within the data analysis, both of which constitute the qualitative research strand that is then triangulated with quantitative data from previous studies.

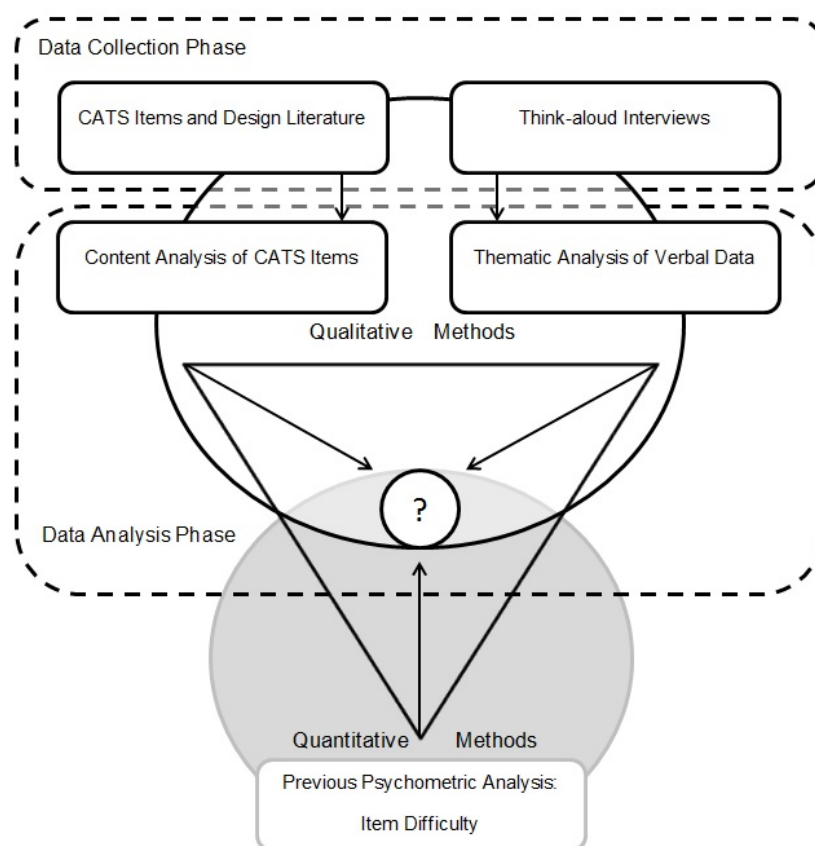


Figure 3.1 Research Design Showing the Qualitative Methods used for Data Collection and Analysis as well as Interpretive Triangulation with Previous Quantitative Analysis

### 3.2.1 Mixing Research Paradigms

Qualitative research is often employed when deep understanding of phenomena is a primary goal. This study not only seeks to develop a deep understanding of difficulty in the context of CATS as a concept inventory, but also seeks to triangulate that deep understanding with psychometric measures that describe the items quantitatively. In order to mix qualitative and quantitative findings in a common analysis, the author as researcher also needed to consider mixing research paradigms.

The interpretivist paradigm assumes that individuals make meaning of their world in diverse and complicated ways. Due to the uniqueness of individual lives, interpretive research seeks to capture aspects of the complex and subjective views that individuals hold regarding the phenomena of interest (Creswell, 2008). In contrast, the postpositivist paradigm assumes a more traditional research worldview in which there are objective, or near objective truths that exist in the world, and through collection of observed evidence, relationships between variables can be determined (Creswell, 2008).

In order to understand the underlying reasoning that accounts for CATS item difficulty, I chose to use interpretive qualitative methods to elicit detail and context for students' responses to CATS items. A key assumption within this perspective includes that due to my experiences as an engineering student and instructor, my participation in the interviews have influenced the analysis. I also assume that I will not be able to capture the whole of student's reasoning, but can use the think-aloud interview as a technique to preserve a snapshot of students' reasoning within the research context. The interpretive perspective extends in some manner to the content analysis of CATS items as



the goal of this analysis is to expand upon the intended target concepts and include alternate possible lines of reasoning that may lead students to select each item response.

In addition to the interpretive perspective, this study also requires a postpositivist perspective in which there is only one objectively correct response to each CATS item. In analyzing students' responses to CATS items, widely agreed-upon scientific principles are used to evaluate deviations from accepted, correct forms of reasoning in order to classify responses as problematic reasoning. Additionally, the psychometric analysis previously performed on CATS response data is also grounded in a postpositivist perspective in which an objective model of student response behavior can be described quantitatively.

The mixing of these research paradigms creates a tension that should be addressed. It is important to acknowledge that this study was situated within different paradigms that influenced research decision-making such as: what constitutes evidence?, or to what degree can inferences based on evidence be made? The most significant decision influenced by this tension is the selection of thematic analysis as the method of analysis. A more detailed description of thematic analysis can be found in chapter section 3.5.2; however the characteristic of thematic analysis that positions findings as a limited perspective on what may be happening with respect to phenomena is particularly appropriate to the blended paradigms at play in this study.

### 3.3 Data Sources

The data of interest for the content analysis strand of the present study are the CATS items selected for analysis and previous literature that was used to code the items based on key aspects of the item development and prior psychometric and validity studies.

The data of interest for the think-aloud interviews strand of the present study are the transcripts of the interviews intended to capture students' reasoning in response to CATS items.

### 3.3.1 CATS

CATS is a twenty-seven item multiple-choice concept inventory that was designed to diagnose students' correct and incorrect understanding of statics concepts (Steif & Dantzler, 2005). The test includes three items for each of nine concepts: (a) Drawing forces on separate bodies, (b) Newton's 3rd Law, (c) Static Equivalence, (d) Roller joint, (e) Pin-in-slot joint, (f) Loads at surfaces with negligible friction, (g) Representing loads at connections, (h) Limits on friction force, and (i) Equilibrium. Each problem was designed to require qualitative reasoning and could be solved without the need for extensive mathematical computation (Steif & Dantzler, 2005). The items were also designed with the intention of diagnosing conceptual errors commonly held by students. These errors include the tendency to leave forces off of a free-body diagram when it should be acting, including forces that should not be acting, and failure to impose a balance of forces in all direction and moments about all axes, among others (Steif & Dantzler, 2005). A detailed description of CATS items and the assessment development of the instrument can be found in chapter section 2.3.4.

### 3.3.2 Source Material for Content Analysis

As described in the literature review, the development of CATS has incorporated numerous stages of item design, implementation and evaluation. The multiple approaches taken to building upon an understanding of CATS as a diagnostic test have produced a wealth of literature presenting findings that may provide additional information into the

context of students' reasoning in response to CATS items. A listing of the publications examined as part of the CATS content analysis is found in Table 3.1; this also includes a brief summary of the types of information relevant to the present study that were identified from each publication.

Table 3.1 CATS Design and Validity Study Literature

Reference	Brief Summary
(Steif, 2003)	Initial conference proceeding on the development of the Statics Concept Inventory with comparison to traditional measures of problem solving ability
(Steif, 2004a)	Initial presentation of underlying conceptual framework, including statics skills and errors
(Steif, 2004b)	Reporting of psychometric findings from initial administration of Statics Concept Inventory (n=125)
(Steif & Dantzler, 2005)	Presentation of refined conceptual framework; reporting of psychometric findings from administration of CI across 5 universities (n=245); factor analysis
(Steif, 2005)	Additional evidence from broader CI administration; detailed analysis of pre/post-test administration
(Steif & Hansen, 2006)	Correlation of CI and course exam performance for multi-institution, revised CI administration (n=1331) [v2004-2005]; Initial identification of common misconceptions
(Hansen & Steif, 2006)	Revised conceptual framework; psychometric analysis of multi-institution, revised CI administration (n=1164) [v2005-2006]; factor analysis of conceptual framework
(Newcomer & Steif, 2006b)	Development of coding scheme and identification of correct conceptions and misconceptions within written explanations of selected CI items (n=39,69)
(Steif & Hansen, 2007)	Initial presentation of distractor coding
(Newcomer & Steif, 2008a)	In-depth analysis of students' ideas of equilibrium as evident in written explanations to a single CI item, #27 (n=58,68)
(Newcomer & Steif, 2008b)	Response pattern coding for written explanations of selected CI items (n=129,128,151,146)
(Santiago-Román, 2009)	Construction, calibration and model fit analysis of the Fusion Model for CATS, a cognitive diagnostic psychometric model consisting of a Q-matrix - a binary representation of underlying cognitive attributes (n=1354) [v2006-2007]
(Santiago-Román, Streveler, Steif, & DiBello, 2010)	Identification of statics cognitive attributes required for each CATS item, development of Q-matrix linking item concepts, concept clusters, and conceptual errors (n=1354) [v2006-2007]
(Santiago-Román, Streveler, & DiBello, 2010)	Identification of mastery profiles – patterns of item and cognitive attribute mastery – based on application of the Fusion Model to CATS (n=1354) [v2006-2007]
(Denick, Santiago-Román, Streveler, & Barrett, 2012)	Pilot think-aloud study used to confirm conceptual framework of CATS (n=5)
(Denick, Santiago-Román, Pellegrino, Streveler, & DiBello, 2013)	Think-aloud study used to confirm conceptual framework of CATS as well as provide evidence of confirmation for the Q-matrix (n=18)
(Jorion et al., 2013)	Psychometric analysis of CATS including test and item performance analysis through CTT and IRT; and structural analysis through factor analysis, subscale reliability and tetrachoric correlations, (n=1372)

Recognizing the need for a diagnostic tool to identify student misconceptions in statics as well as inform instruction, Steif sought to establish a conceptual framework for statics and develop the Statics Concept Inventory, later referred to as the Concept Assessment Tool for Statics (CATS). The CATS instrument may be viewed as further along in its realization than other engineering CI's, as indicated by a published psychometric analysis of test items which includes an analysis of demographic variance; this analysis also used Item Response Theory (IRT) to evaluate the instrument and provide *difficulty* and *discrimination* indices for each item. Additionally, CATS was evaluated for reliability via *Cronbach's alpha* and validity was established through evidence of *content validity*, *criterion-related validity*, and *construct validity* (Steif & Dantzler, 2005). Further administration of the instrument has led to significant findings into students' common errors and misconceptions (Steif & Hansen, 2006).

Based on the most recent and extensive psychometric evaluation of CATS (Jorion et al., 2013), the following provides an overview of the quantitative interpretive argument for CATS:

- Item difficulties ranged from 0.25 to 0.78, with the exception of item 26 at 0.16. This is the first indication of the problematic performance of item 26. This also falls within the acceptable range of 0.2 – 0.8.
- Item discrimination ranged from 0.20 to 0.49, with the exception of item 26 at 0.18. These values are reasonable and equal to or greater than the recommended value of 0.20.
- After removing problematic item 26, Cronbach's alpha was 0.86 indicating good reliability for a formative assessment

- Four approaches to evaluate the conceptual structure of CATS: subscale reliability, tetrachoric correlations, exploratory factor analysis, and confirmatory factor analysis closely aligned with the intended conceptual structure of CATS with few exceptions, for details on the structural analysis of CATS see (Jorion et al., 2013)
- Diagnostic modeling of CATS examined the extent to which diagnostic outcomes for students and groups of students are statistically supported. CATS showed strong diagnostic strength relative to identified cognitive attributes, for details on Q-matrix development and application of the Fusion Model see (Santiago-Román, 2009).

### 3.3.3 Interview Participants

In contrast to previous psychometric studies that were based on analyses of large data sets collected from CATS administrations in undergraduate engineering classes, the present study builds upon previous item analysis and validity studies by incorporating detailed verbal response data obtained through think-aloud interviews in a clinical setting. Verbal data may provide additional insight into students' reasoning in response to CATS items as the verbal data collected through think-aloud interviews may describe students' reasoning more richly and in greater detail. Semi-structured think-aloud interviews were conducted in April of 2012; IRB approval was obtained prior to conducting the student interviews. The participants in this study were 18 undergraduate engineering students from a large, public Midwestern university. Although not intended to provide a representational sample, students were recruited with regard to engineering coursework progression so that the sample would include students at a similar academic stage as those that would likely encounter CATS in the classroom. All of the students were in

their second or third year and had completed a statics course within one year of the interviews. The sample consisted of 13 males and 5 females and included students majoring in mechanical, civil, and industrial engineering.

The sample of 18 undergraduate engineering students was obtained through email recruitment. Using departmental mailing lists, emails were sent to undergraduate students within aerospace, mechanical, civil, and industrial engineering departments, as these departments are likely to house students who had completed a statics course. In addition, recruitment emails were sent to recipients of a women in engineering program and a minority engineering program. This may explain the relatively high proportion of female students in the sample. To participate in the study, the students were required to have completed a statics course within the previous year and have the ability to explain their thinking process in fluent English when solving a problem. Eligible students were then interviewed and compensated for their participation; each interview was completed within a two-hour window.

It is important to note that as the researcher who conducted the interviews and analyzed the verbal data, I hold an undergraduate degree in mechanical engineering and have experience teaching statics and introductory mechanics. This background allowed me to approach CATS with an expert-like perspective and was useful in determining the meaning of student explanations and any deviation from accepted scientific reasoning of specific statics concepts.

### 3.4 Data Collection

#### 3.4.1 Content Analysis Data Collection

A summary of the literature collected for analysis as part of the content analysis research strand can be found in chapter section 3.3.2.

#### 3.4.2 Think-aloud Interviews

A common method for collecting cognitive data is the think-aloud interview method that shares a key characteristic with verbal protocol, protocol analysis, or verbal analysis in that verbal data is collected as a primary form of evidence (Chi, 1997; Ericsson & Simon, 1993). Protocol analysis, as described by Ericsson and Simon (1993), is both an empirical method and a theoretical approach, based on the caveat that the cognitive processes that generate verbal data can provide evidence that allows for inference of internal reasoning. This form of research relies heavily on external representations, including verbalizations and drawings to form an understanding of the reasoning used in cognitive tasks (Chi, 1997). The think-aloud interviews conducted in this study were approved by the Purdue Institutional Review Board; IRB approval documentation can be found in Appendix A.

#### 3.4.3 Interview Materials

Two booklets containing 8 CATS items each were created in order to collect verbal explanations for a total of 14 CATS items. The selected items were chosen for specific combinations of skills and errors, with a breadth of item difficulty. The items increased in difficulty across each booklet with some separation between items that addressed similar concepts.

#### 3.4.4 Interview Protocol

The interview protocol was informed by findings from a previous pilot study (Denick et al., 2012). As students' reasoning and thinking cannot be determined solely from item responses, I prompted students to explain their thinking for individual CATS items and to describe why they did not select alternate responses. Students were encouraged to verbally explain their thinking as they initially approached each problem and after arriving at an answer, students were prompted for further explanation regarding specific aspects of the problem and why they did not select alternate responses. I opted to question students iteratively by returning to previously answered problems after addressing all of the items in the booklet. Further prompts regarding interpretation of problem statement and diagrams, and specific aspects of student reasoning were posed to students to allow for multiple modes of student explanation. Audio recordings were taken at the time of the interviews and transcripts were created to analyze students' thinking. The interview protocol used in this study can be found in Appendix B.

#### 3.4.5 Use of Existent Data

It is important to note that the interview data analyzed in the present study was previously collected for qualitative analysis as part of Pellegrino and DiBello's research on a comprehensive approach to validity for CIs in engineering education (Denick, Santiago-Román, Pellegrino, et al., 2013; Denick, Santiago-Román, & Streveler, 2013; Denick et al., 2012). The interviews conducted for the previous qualitative analysis intended to elicit students' reasoning about key statics concepts as a means of collecting confirmatory evidence for an existing conceptual framework.



The use of existing data in research has an obvious advantage in that a large data set can be obtained in a short amount of time. The major disadvantages of using existing data may be described as limited control over the protocol used to collect data, and problems with validity. The use of existing data limits the overall flexibility of the research design, as an alignment of inquiry should be maintained between research purpose, questions, data collection and analysis (Creswell, 2008).

As mentioned, the think-aloud interviews were conducted as part of a previous verbal protocol study. The purpose of the previous study was primarily to investigate the alignment of students' reasoning in response to CATS items with the skills and errors comprising the CATS Q-matrix. The interview data collected for the previous validity study is appropriate for use in the current extended validity study due to the focus on eliciting student reasoning. For both studies, the extent of the rich description found in students' explanations of their reasoning is of paramount importance. In the case of the previous validity study, an a priori coding scheme allowed for examining to what extent the detailed evidence supported a theoretical model of skills and errors argued to underlie CATS. In this study, the detailed interview data was used to develop an argument that would provide insight into a related psychometric construct, *item difficulty*.

### 3.5 Method of Analysis

#### 3.5.1 CATS Item Grouping

Previous analysis of data collected from multi-institution CATS administration have determined psychometric properties of CATS items, including item difficulty values calculated using methods associated with classical test theory and item response theory (Table 3.2) (Jorion et al., 2013; Steif & Dantzler, 2005)

Table 3.2 CATS Item Difficulty as Determined through Classical Test Theory, and 1- and 2-Parameter Item Response Theory

CATS Item	Classical Test Theory Item Parameters		Item Response Theory 1PL Model Item Parameter	Item Response Theory 2PL Model Item Parameters	
	Item Discrimination	Item Difficulty	Item Difficulty	Item Discrimination	Item Difficulty
1	.483	.513	-0.21	1.56	-0.08
3	.472	.585	-0.57	1.62	-0.34
4	.440	.254	1.23	1.26	1.10
5	.316	.437	0.18	0.79	0.35
7	.427	.316	0.84	1.22	0.80
8	.406	.351	0.64	1.12	0.67
14	.444	.697	-1.17	1.58	-0.78
15	.399	.735	-1.40	1.42	-0.99
17	.397	.264	1.16	1.12	1.13
18	.361	.576	-0.52	0.97	-0.39
19	.395	.626	-0.78	1.13	-0.59
22	.312	.292	0.98	0.79	1.26
23	.386	.286	1.02	1.04	1.06
27	.396	.487	-0.07	1.06	0.04

Additionally, using a Q-matrix approach cognitive attributes were identified that cut across items as pieces of domain understanding that students would need in order to correctly answer specific CATS items (Santiago-Román, Streveler, Steif, et al., 2010). Using a large set of student responses, Santiago-Román was also able to determine which cognitive attributes were more difficult. Results from an analysis of mastery profiles revealed that the following cognitive attributes have the highest occurrence of non-mastery: friction force, Newton's 3rd Law, contact forces, couples and equilibrium, and equivalence (Table 3.3) (Santiago-Román, Streveler, & DiBello, 2010).

Table 3.3 Problematic Cognitive Attributes as Determined by the Fusion Model

Cognitive Attribute	Name	Items
1	Equivalence	7, 8, 9
2	Newton's 3 <sup>rd</sup> Law	4, 5, 6
5	Contact Forces	2, 16, 17, 18
7	Friction Force	22, 23, 24
21	Couples	7, 8, 9, 25, 26, 27
13	Equilibrium	25, 26, 27

Based on this data, the CATS items selected for the current study were grouped into three cases: more difficult items, less difficult items, and items with difficult cognitive attributes (Table 3.4).

Table 3.4 Case Groups by Difficulty

Case Groups	CATS Items	Concept Grouping (Steif and Hansen, 2007)
Less Difficult Items	1 & 3	Free Body Diagrams
	14 & 15	Pin-in-Slot
	19	Representation
More Difficult Items	4 & 5	Newton's 3 <sup>rd</sup> Law
	7 & 8	Static Equivalence
	22 & 23	Limits on Friction Force
Items with Difficult Cognitive Attributes	17 & 18	Negligible Friction
	27	Equilibrium

### 3.5.2 Content Analysis

In developing CATS, Steif drew upon his and others' previous research regarding key concepts and misconceptions that students demonstrate in reasoning about statics problems (Steif & Dantzler, 2005). In order to better understand the development of CATS items, I opted to perform a content analysis by examining all descriptions and analysis of the CATS items used in the think-aloud interviews across all the CATS design

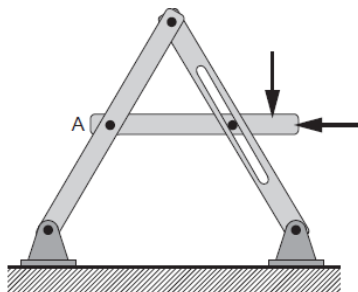
and development literature. Over the course of the analysis research stage, I met with Paul Steif on three occasions for conversations lasting 1-2 hours to discuss the data collected through the content analysis and to fill in any unknown design intentions. Because I was seeking to gain a better understanding of the CATS items as a contextual element, and to expand upon possible forms of desired and problematic reasoning this approach to content analysis aligns well with Krippendorff's (2013) description of qualitative content analysis within a larger qualitative study. As an example of the data collected in the analysis, the following includes a sample analysis of Item 5 as shown in Figure 3.2.

**Items 4 & 5**

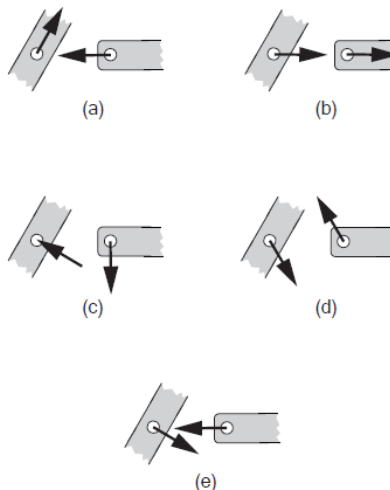
**CMU Coding** (Hansen & Steif, 2006): Concept B: Newton's 3<sup>rd</sup> Law; Concept Cluster: 1, 4

**Q-matrix Coding** (Santiago-Román, Streveler, Steif, et al., 2010): Cognitive Attribute 2: Newton's 3<sup>rd</sup> Law; Expected Common Errors: 2, 6, 8

5. The frame shown could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin at A.



Which of the following could represent the forces exerted by the pin on the two bars? Forces act in the senses shown.



**Misconceptions** (Steif & Hansen, 2007): Force must be parallel to member (E6), Force must be perpendicular to member (E6)

**Distractors** (Steif & Hansen, 2007):

5.     A: Force must be parallel to member
- B: Null (*Possible problematic substance-based reasoning, transfer of force from one member to another*)
- C: Force must be perpendicular to member
- D: Correct
- E: Null

**Desired Reasoning:** Students are expected to recognize that when separating a body for analysis, any force pairs at a connection act in equal and opposite directions.

Students need to understand that the forces shown in the given diagram are representative of any forces that may be acting on a body and not the specific forces of interest for analysis.

Figure 3.2. CATS Item 5, Targeted Concept of Newton's 3rd Law

Item 5 was designed to assess students' understanding of Newton's 3rd Law, that "forces between two contacting bodies must be equal and opposite (Steif & Dantzler, 2005)." Incorrect responses to this item may also indicate evidence of the following errors in students' reasoning: "failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two.", "failing to account for the mutual (equal and opposite) nature of forces between connected bodies that are separated for analysis", and "not allowing for the full range of possible forces between connecting bodies, or not sufficiently restricting the possible forces (Steif & Dantzler, 2005)." In connecting this item to later work, I was able to determine that the item was coded for "Newton's 3rd Law" as both a concept and a cognitive attribute.

In addition to cross-referencing the items across the CATS development literature, I also chose to explore alternate interpretations that students may use in response to each of the items. In this case, I noted that in addition to the targeted concept that forces between objects must be equal and opposite, students may also need to recognize that when separating a body for analysis, any force pairs at a connection act in equal and opposite directions and the students may also need to understand that the forces shown in the given diagram are representative of any forces that may be acting on the system and are not necessarily specific forces of interest for analysis.

Through conversations with Paul Steif about the design of CATS items, additional resources were identified to add to the content analysis, namely a record of the specific misconceptions or errors associated with the distractors for each item. Most CATS item distractors were developed through open-ended student interviews in which students

responded to the item problem stems without multiple choice response options. Common incorrect responses to the open-ended versions of the CATS items were developed into distractors. At a later point, the distractors produced from the interviews were evaluated to identify likely errors in reasoning that would lead students to each incorrect response. The reasoning provided for each distractor through this initial analysis is identified in this study as expected reasoning. However it is important to clarify that the expected reasoning identified for each distractor does not fully encompass the CATS distractor design or targeted misconceptions, but serves as an initial attempt to label possible explanations for why students commit the common errors contained in the distractors as found through open-ended student interviews.

Item 5 includes two distractors, A and C, which may align with misconceptions that “force must be parallel to member” and “force must be perpendicular to member” respectively. These misconceptions aligning with distractors A and C for Item 5 were assumed to be the result of students’ confusing the force conditions of two-force members. Two-force members are common structural elements in which forces act at each end of a member, without a resultant couple. This requires that under equilibrium conditions, the two forces acting at the ends of the member are equal and opposite of one another. Although not considered in the initial analysis of the item or its distractors, another possible form of problematic reasoning that may be triggered by this item relates to naïve physics reasoning in which force is thought of as a substance that is passed from one object to another (Reiner et al., 2000). Distractor B may potentially align with this problematic reasoning as the forces in the same direction may be misconstrued as the force in one member transferred to the other member.

While not all of the CATS items of interest could be located in all of the CATS development literature, attention was paid to perform the content analysis in as systematic a manner as possible. Because the content analysis informed and was informed by the simultaneous thematic analysis, it was appropriate to use an interpretive approach to the content analysis and describe intrinsic and extrinsic properties of the data as opposed to using frequency counts of word usage or more quantitative analytical methods (Krippendorff, 2013). In summary, each of the CATS items were cross-referenced with item descriptions included in CATS development literature, additional design considerations were obtained through discussion with the assessment designer, and alternate forms of reasoning or cognitive processes were proposed that may be triggered by the items both as desirable and problematic forms of reasoning. The content analysis for each CATS item of interest for this study can be found in Appendix C.

### 3.5.3 Thematic Analysis

Although presented as separate qualitative research strands, it is important to note that the content analysis of CATS items and the thematic analysis of the think-aloud interview data were performed simultaneously and informed each other throughout the data analysis stage of the present study. Similar to content analysis as a method, thematic analysis can be broadly described as a systematic approach to encoding qualitative data (Boyatzis, 1998). Thematic analysis may be used to categorize data, develop themes and interpret aspects of a research topic among other purposes (Boyatzis, 1998; Braun & Clarke, 2006). This is a highly flexible approach to analysis of qualitative information that often serves as a basis for other qualitative methods. Thematic analysis may be distinguished from similar approaches such as grounded theory in that the approach does



not adhere to a single epistemological perspective and does not require the development of a formal theory (Braun & Clarke, 2006; Guest, MacQueen, & Namey, 2012).

Thematic analysis methods were appropriate for this study due to the type of data available and collected, and the level of insight expected. Think-aloud interviews with only a small sample of undergraduate engineering students can only provide some indication of the possible myriad of divergent reasoning that may exist as students respond to CATS item. The nature of the study to only provide a limited perspective into students' reasoning made thematic analysis a more appropriate approach than grounded theory. Phenomenological methods would also not have been appropriate as the study incorporates a postpositivist perspective in that there is judgment of correct and incorrect reasoning in response to CATS items and was not solely interested in exploring students' experiences interacting with CATS items.

An inductive coding approach was taken as a means of identifying themes across data sets. This approach may also be described as exploratory or content-driven in that the codes were not predetermined and derived from the data (Guest et al., 2012). Rather than simply counting the frequency of words or phrases, the thematic analysis used for this study involved describing both the intrinsic and extrinsic ideas that could be drawn from a deep analysis of the data.

In a previous study examining student responses to CATS items, Chi's (1997) method of verbal data analysis was used to inform the analytic approach (Denick, Santiago-Román, & Streveler, 2013). For this study, I opted to employ the recursive approach to thematic analysis as described by Braun & Clarke (2006) within an initial approach to develop a thematic scheme before applying Chi's (1997) approach to

systematically analyzing the instances of the codes obtained after the themes were established. The previous study based on students' think-aloud responses to CATS items sought to examine to what extent student reasoning aligned with an a priori coding structure of skills and errors, and in that case it was more useful to immediately quantify forms of verbal response. In the present study, I hoped to explore the interaction between CATS item design and students' reasoning and sought to begin with a more descriptive analysis before quantifying the qualitative themes to interpret my analysis across item case groups.

The following series of figures (Figures 3.3-3.5) shows the progression of thematic formalism that was explored through the thematic analysis. A full description of the final themes and codes is included in chapter section 3.6.

### Progression of Thematic Scheme

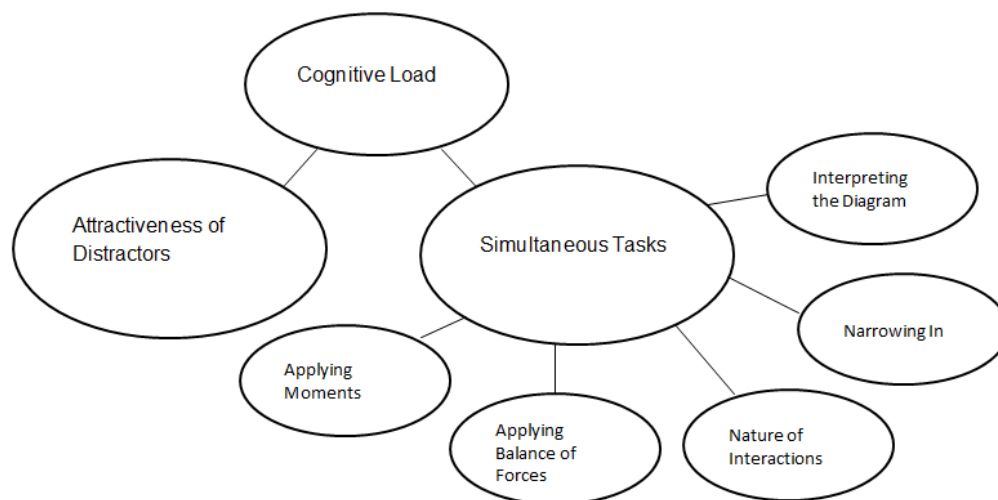


Figure 3.3. Initial Thematic Map, Showing Two Main Themes

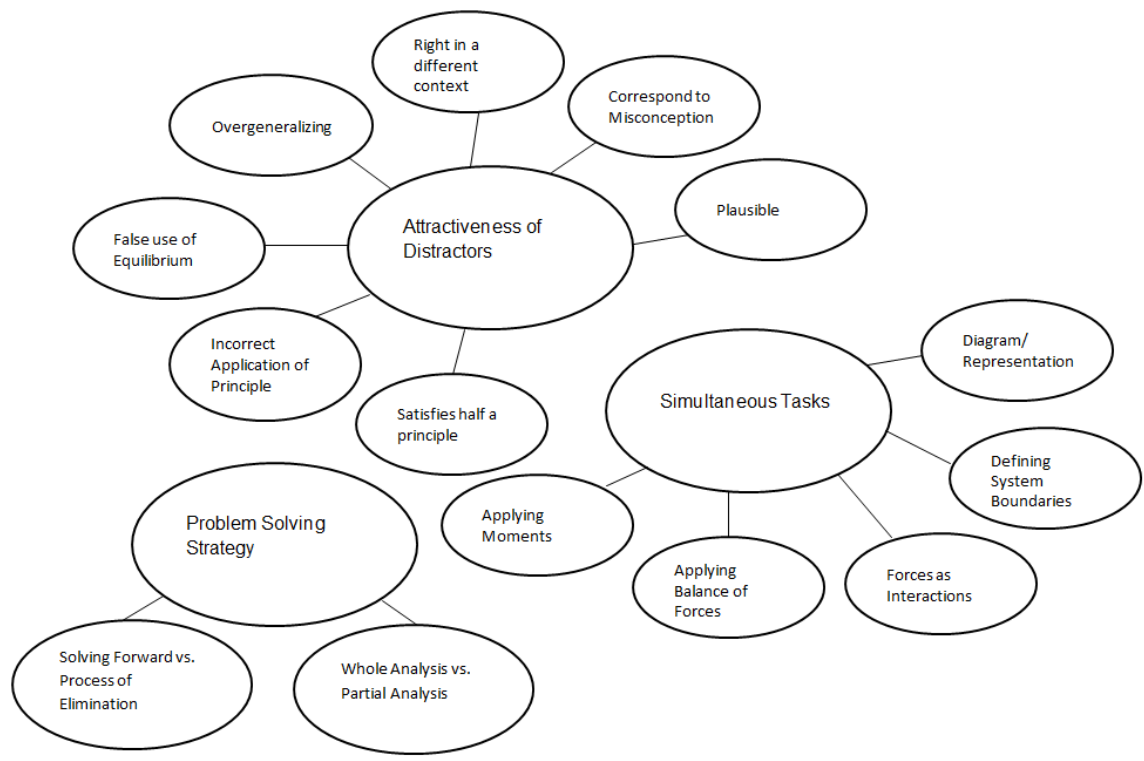


Figure 3.4. Developing Thematic Map, Showing Three Main Themes

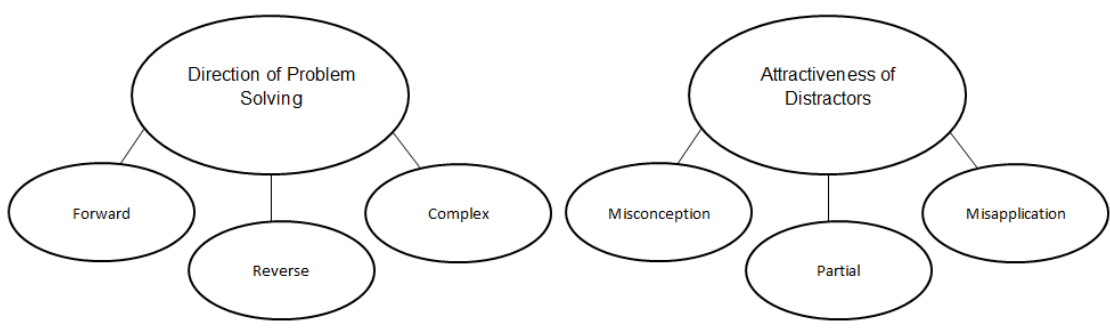


Figure 3.5. Final Thematic Map, Showing Two Main Themes

Once the thematic analysis yielded a stable thematic structure as shown in Figure 3.5, the analysis incorporated additional analytic steps following Chi's (1997) approach to verbal analysis. This approach was taken as the purpose of the study aligns with the stated goal of verbal analysis, "the goal of the method here is to attempt to figure out

what a learner knows (on the basis of what a learner says, does, or manifests in some way, such as pointing or gesturing) and how that knowledge influences the way the learner reasons and solves problems, whether correctly or incorrectly. The following steps as adopted from Chi's recommendations describe the approach taken to code and analyze the verbal data collected in response to CATS items:

1. Developing or choosing a coding scheme or formalism
2. Operationalizing evidence in the coded protocols
3. Depicting the mapped formalism
4. Seeking patterns in the mapped formalism
5. Interpreting the patterns

Although Chi (1997) describes 8 functional steps of verbal analysis, not all of the steps of the cited approach were applicable to this work. An initial reduction or sampling of protocols is recommended, however in this study, the sampling approach yielded a manageable amount of transcript data. Similarly, a recommendation for a final step of repeating the process at a different granularity of analysis was omitted, since sufficient findings were obtained at the selected level of analysis. The removal of these steps do not alter the approach, as some flexibility is inherent to verbal analysis as described by Chi, and some steps are considered optional to the process.

### 3.6 Theme Definitions

The following sections, 3.6.1 and 3.6.2, describe the two main themes and associated codes as shown in Figure 3.5, including examples of evidence from the

content analysis and think-aloud interviews that provide reasoning for the assignment of specific codes.

### 3.6.1 Direction of Problem Solving

The 'Direction of Problem Solving' theme looks at the intersection of the design of CATS items with respect to the type of multiple choice formats, the direction of the solution strategy that students take when responding these items, and the difficulty of the items as determined from previous studies. The coding for this theme emerged from thematic analysis of think-aloud interview data in concert with content analysis of the items themselves. These codes were built upon previously identified typologies of multiple-choice items (Case & Swanson, 2001; Haladyna, 2012).

The following three codes describe the types of multiple choice items included in CATS:

- **[Forward]** A-type/Conventional: a single, correct response can be identified from the problem stem. In this case, it is expected that students will solve problems in a forward direction by reading the problem stem, identifying a response and then finding a matching response from the multiple choices.

Example: What is the capital of Spain?

- a) Barcelona
- b) Madrid
- c) Seville
- d) Valencia

- **[Reverse]** A-type/Conventional: a single, best response can be identified by evaluating the distractors. In this case, students must work backwards from the responses to see which option correctly responds to the given problem stem.

Example: Which of the following cities is a state capital?

- a) Baltimore
- b) Los Angeles
- c) Baton Rouge
- d) Las Vegas

- **[Complex]** K-type multiple choice items: a primary set of responses is presented, and the applicability of those responses evaluated by a secondary set of possible combinations. In this case, student may work forward

Example: Which of the following states borders Canada?

- |                   |                |
|-------------------|----------------|
| a) North Carolina | I. a & b only  |
| b) North Dakota   | II. a, b, & c  |
| c) New Hampshire  | III. b, c, & d |
| d) New Mexico     | IV. b & c only |

Items coded to both the Forward and Reverse codes are considered Conventional (Haladyna, 2012) or A-type, One-best-Answer (Case & Swanson, 2001) multiple choice items. These are the most common form of multiple choice items and a wealth of research exists that examines aspects of this multiple choice format. I have chosen to use a directionality convention to look at a perceived difference between these forms of conventional multiple choice items. This was chosen to align with the differences among

how students reasoned through these seemingly different forms of CATS items. Another way to think about what makes Forward and Reverse items different is that with Forward items, a clear single solution can be determined from the information given in the problem stem whereas with a Reverse item, a number of possible scenarios may result in a correct response to the problem stem, however only one of the responses provided is correct. Other literature refers to this distinction as a Recall item versus a Best Choice or Application item (Case & Swanson, 2001), referencing a distinction in the level of reasoning required similar to Bloom's Taxonomy.

Although not differentiated in the cited references, the key distinction between multiple choice items coded [Forward] and [Reverse] may be also described as determinant and indeterminant. A directional convention was selected as an easy way to grasp the different approaches that are needed to solve these types of multiple choice items.

In deciding how to code the interview transcripts with respect to the direction of problem solving, only students' initial response to the item prompt were included. The interviews were conducted in an iterative manner where students were asked to respond to the items multiple times as a means of obtaining rich data about students' reasoning in response to the items. However, in this case, the primary focus is on how students responded during the first iteration of CATS item prompting. Because students often started with one approach and then abandoned that line of reasoning for another, the coding for a single student's response to an item may include a series of directional codes.

- **[CF] Correct Forward:** Student uses the information from the problem stem to arrive at a solution. This solution is then checked against the provided responses for a match. A correct response is selected.

**[CF]** Item 3, Student Q: *“And for this, I'd draw a free-body diagram, and we want to talk about 1, 3 and 6. So we need this one, this one and this one. And I'm assuming, it tells me that each one has its own weight. And each cord has its own tension. So the correct free-body diagram we'd have-- I'm just going to try to do this without looking at the responses and then I'll pick one-- whichever one is the closest on my free-body diagram. So first, we're going to have-- just talking about Block 1, we're going to have the weight of Block 1 and we're going to have the tension from A. And we also want Blocks-- I'll just go ahead and draw 3, 6. All right. And then on the left side, we are-- they're not drawn, but each one has-- each block is pulling down, Blocks 2, 4 and 5 are pulling down on 1 from the left end. So we're going to have the tension of B. And then we're going to have-- we're going to neglect the tension of D and E. And so we've just got the tension of B, and then we've got the weight. We've also got the weight of 2. And then with 2, includes a weight of 4, and the weight of 5. Block C, Block 3, is going to look the same. And there's going to be-- there's not going to be a tension between 1 and 3 because equal and opposite. So I'm going to eliminate the tension of C. Also eliminate the tension of F. Block 3 is going to have a weight. Block 6 is going to have a weight. And then there's going to be tension and a weight from 7. Tension G and on 7. Now I'm going to go back and look between my choices, and pick the one that's closest to what I drew... So my final choice is C.”*

- **[IF] Incorrect Forward:** Student uses the information from the problem stem to arrive at a solution. This solution is then checked against the provided responses for a match. An incorrect distractor is selected.

**[IF]** Item 23, Student V: *“So, I guess all I'm looking for is a frictional force. So, friction is normal times the coefficient. And I can just disregard this 10 newton force. That's not needed. So, in this case, the normal is going to be the total of the weights. So, that's 90 newtons. And the coefficient of friction is point two. And so, if I were to do that, then I get-- so 90 divided by-- is going to be 18, E. Yeah that's my answer. ... I guess this one is more of just a simple equation. I guess I'm also assuming that--*



*well, no I think-- I think it's just a straightforward equation. That's all you need to know."*

- **[NF] Null Forward:** Student expects to work from the problem stem and arrive at a single correct solution, but is unable to generate solution.
- **[CR] Correct Reverse:** Student evaluates the responses with respect to the problem conditions. A correct response is selected.

**[NF -> CR]** Item 3, Student V: *"So, first I'll begin by drawing what I think it will be. ... Yeah. So, for the first block 1, I'm going to first draw the weight, which is acting through the center, and then the tension, which counteracts that. And then on the side, on side C, there would be a tension going up, a weight going down, and then two tensions acting down. And then, on block 6, there's a weight going down, a tension going up. And the same with 7. And this matches--*

*\_\_\_\_\_ . Okay, so it includes blocks 1, 3, and 6. So, everything else would be external to the system. So, there's also a tension here that includes the weights of blocks 2, 4, and 5. So, it's just the tension of B is what I'm going to say. And then on the other side, the weight of 3 would be acting downwards and the tension of G, which isn't in the system and then weight of 6. So that most closely matches answer C for me, actually. So, I guess the way I solved this was by looking at the answers, and then going back and figuring it out."*

- **[IR] Incorrect Reverse:** Student evaluates the responses with respect to the problem conditions. An incorrect distractor is selected.

**[IR]** Item 15, Student S: *"Okay, so as you can see by the rod which has the weight attached to it, it's going to move down, because it has a weight attached to it, and the force acting downwards. So the angle would drop as the object moves down. So the force exerted by the slot should be on the bottom part of the slot, of the hole in the steel rod. No, I'm sorry. The force should be on the top part of the steel, the hole in the steel rod, as it goes down. The pin hits the roof of the hole. And so it would be from the top. And so that rules out Answer D and C. And so the remaining Answer A, B and E. And the force would not be straight down, because the steel rod is at an angle, so that rules out Answer A. So the only possible answers are B and E. And since it's going to move down, there's going to be a rotational*

*momentum to it. Or rotational motion. So B is the only answer with rotational motion, so the answer would be B.”*

- **[NR] Null Reverse:** Student evaluates the responses with respect to the problem conditions. The student is unable to arrive as a single response. This code did not appear in any student interview transcripts.

### 3.6.2 Distractor Attractiveness

The following codes were developed to describe concept inventory distractors and why students may incorrectly select specific distractors. These codes emerged from iterative thematic analysis of a content analysis of CATS design literature, and verbal analysis of students' response to CATS items in think-aloud interviews. Both of these approaches allow for slightly different insight into how the design of the CATS distractors interacts with the student reasoning behind incorrect responses and selection of specific distractors.

Previous studies on CATS have described the alignment of the distractors (Steif, 2004; Steif & Dantzler, 2005; Steif & Hansen, 2007) with known misconceptions and common errors that students make in Statics. Through analysis of the concept inventory itself, reference to previous literature and discussion with CATS designer Paul Steif, a content analysis was performed that identified additional characteristics of the item distractors. Using these resources as source materials, the following codes were applied to the reasoning described in the test development documentation to identify the expected reasoning of each distractor. The coding applied to the student interview data was identified as the observed reasoning for each distractor.

The following three codes describe different ways that students may incorrectly respond to a CATS item and select any given distractor:

- **[Misconception]** Corresponds to common misconception: distractors that align with established misconceptions from physics, mechanics, etc...
  - Examples: Force along a member (substance-based reasoning); rotation as evidence of a moment (similar to a rotational impetus theory)
- **[Misapplication]** Inappropriate application of principle: distractors that align with the misapplication of a principle or when an equation or rule is applied to an incorrect context
  - Examples: Force acts to balance an external force (misapplication of Equilibrium); Moment/couple acts due to a force at a distance (misapplication of principle of moments); Friction force must equal  $\mu N$  (misapplication of static friction formula)
- **[Partial]** Satisfies principle partially: distractors that satisfy one aspect of a principle, but do not satisfy the principle entirely
  - Examples: Forces balanced, but not moments; Moments balanced, but not forces (partial application of equilibrium conditions)

A detailed discussion of how these codes were applied to student responses can be found in chapter section 4.3.

### 3.7 Quality Considerations

There are multiple, varying viewpoints on how qualitative research should be evaluated with respect to quality (Creswell, 2008). Validity and reliability may not be

appropriate indicators since paradigms of qualitative research do not use evidence in the same manner that traditional scientific research does. This dissertation study in particular is difficult to assign to a specific framework for quality because it mixes interpretivist and postpositivist paradigms. Considering this, I have opted to summarize the key aspects of the method that may serve as indicators of quality. Firstly, the data collected for this study adhered closely to recommended strategies of content analysis and think-aloud interview. As evident in the discussion on data sources, close attention was paid to appropriate selection of materials and appropriate population sampling for the purposes of the study. Secondly, throughout the analysis stage of this dissertation study, particular attention was paid to documentation of theme development and adherence to established strategies for conducting thematic analysis. Finally, participation in research group meetings with experience educational researchers informed the selection and development of interview materials and guidance in performing data collection activities and extensive discussions were held with the designer of CATS as a means of mitigating error in describing design intentions as well as ensuring accuracy in scientific reasoning and inferences.

## CHAPTER 4. FINDINGS

### 4.1 Introduction

The purpose of this dissertation study is to provide context for why some CATS items are more difficult than others. Specifically, the research question guiding this study is:

- How does student reasoning in response to CATS items explain variance in item difficulty across test items?

Through a multi-strand qualitative research design, the following themes emerged as possible explanations for why some CATS items are more difficult than others: (1) a Direction of Problem Solving theme describes the direction of reasoning required or used to respond to CATS items, and may also provide some description of students' reasoning in response to determinant and indeterminant items; and (2) a Distractor Attractiveness theme describes problematic reasoning that is expected and observed in incorrect CATS responses. The following presentation of findings includes summaries and descriptions of thematic coding and comparisons across groups; discussions of the meanings of these findings are included in chapter section 5.2 of this document.

### 4.2 Direction of Problem Solving

Although all multiple choice items present test takers with a problem and a set of responses among which to select a correct response, not all multiple choice items require

the same forms of reasoning to solve a problem. The Direction of Problem Solving theme includes two coding levels: item-level codes that describe the direction of reasoning aligned with the design of the item, and student response-level codes that describe the directions or problem solving approaches that students' take in response to CATS items.

#### 4.2.1 Direction of Problem Solving at the Item-Level

The Direction of Problem Solving theme as applied at the item-level consists of three codes: (1) Forward, which indicates items that can be solved directly from the problem stem and do not require evaluation of all responses to determine a correct response, (2) Reverse, which indicates items that require a comparison among responses in order to find a correct response, and (3) Complex, in which a primary set of responses is provided and students must select from a second set of responses that evaluates the primary responses in a variety of combinations. Table 4.1 includes the item-level codes for each CATS item group.

Table 4.1 Item-level Coding within Direction of Problem Solving Theme

CATS Item Groups	Cats Item and Associated Direction					
Less Difficult Items	1	3	14	15	19	
	Forward	Forward	Forward	Forward	Forward	Forward
More Difficult Items	4	5	7	8	22	23
	Reverse	Reverse	Reverse	Reverse	Forward	Forward
Items with Difficult Cognitive Attributes	17		18		27	
	Complex		Complex		Reverse	

It appears that less difficult items have an association with a Forward direction of problem solving where a correct response can be determined solely from the problem stem without requiring evaluation of all provided responses. Items that require

evaluation of all responses to find a correct solution, coded as Reverse, tend to be associated with more difficult CATS items; these items may also be described as including indeterminant problem stems.

#### 4.2.2 Direction of Problem Solving at the Student Response-level

In addition to item-level coding, the direction of students' reasoning as evident in verbal data was coded at the student response-level. Because students often started with one problem solving approach and then abandoned that line of reasoning for another, the coding for a single student's response to individual items may include a series of directional codes. Table 4.2 includes the student response-level codes for the less difficult CATS items that indicate both direction and correctness of reasoning.

Table 4.2 Student Response-level Coding within Direction of Problem Solving Theme,  
Less Difficult CATS Items

Student Group		Less Difficult Items				
Item	1	3	14	15	19	
Direction	Forward	Forward	Forward	Forward	Forward	
<b>Booklet 1</b>	K	CR	CF	CF	CF	
	M	CF	CF	CF	CF	
	P	CF	CR	CF	IF -> CR	
	R	NF -> CR	NF -> CR	CF	CF	
	S	IR	IR	NF -> CR	IR	
	T	CR	CR	CF -> IR	CF	
	U	CF	CF	CF	IF	
	V	CF	NF -> CR	CF	CF -> IR	
	W	CR	CF	CF	CF	
<b>Booklet 2</b>	J		CF	NF -> IR		IR
	L		CF	CF -> IR		IF
	N		CR	CR		CF
	O		IR	IR		IR
	Q		CF	CR		IR
	X		CF	CF		CF
	Y		CR	CF		CR
	Z		IF -> CR	CF		CR
	AA		CR	NF -> CR		CR

Tables 4.3 and 4.4 include the student response-level codes for the more difficult CATS items and items with difficult cognitive attributes indicating both direction and correctness of reasoning. Note that since students were presented with different items depending on the booklet used for each interview, gray cells indicate items that a student did not respond to.



Table 4.3 Student Response-level Coding within Direction of Problem Solving Theme,  
More Difficult CATS Items

Student Group		More Difficult Items						Student Group
Item	4	5	7	8	22	23	Item	
Direction	Reverse	Reverse	Reverse	Reverse	Forward	Forward	Direction	
<b>Booklet 2</b>	J	CR	NF -> CR	CR	CR	NF -> IR	NF -> CR	<b>Booklet 1</b>
	L	IF	NF -> CR	CR	IF	CF	CF	
	N	IF	NF -> IR	IR	IR	IF	IF	
	O	IF	NF -> IR	IR	IF -> IR	CF	IF	
	Q	NF -> IR	CF -> IR	IR	NR -> IR	NF -> IF	IF	
	X	NF -> CR	NF -> CR	NR -> IR	IF -> CR	NF -> IR	NF -> CF	
	Y	NF -> CR	CR	CR	IR -> CR	NF -> NR	NF -> IF	
	Z	CR	NF -> CR	IR	IR	IF	IF	
	A	NR -> IR	CR	CR	CR	NF -> IF	IF	
	A							

Table 4.4 Student Response-level Coding within Direction of Problem Solving Theme,  
CATS Items with Difficult Cognitive Attributes

Student Group	Items with Difficult Cognitive Attributes			
Item	17	18	27	
Direction	Complex	Complex	Reverse	
<b>Booklet 2</b>	J	CF -> IR	CF -> IR	IR
	L	IR -> CR	CR -> IF	CR
	N	IR	CR	IR
	O	IR	NF	IR
	Q	IR -> NF -> IR	NF -> IR -> NF	IR
	X	NF -> IR	CR	NF -> IR
	Y	IR	CR	IF
	Z	CF -> CR	CR	CR
	AA	CR	CR	CR

Using the student response-level coding as a basis for comparing the less difficult CATS items and the more difficult CATS items, students showed evidence of changing their problem solving approach nearly twice as often in response to more difficult items.

Additionally, although only evident in Student Q's responses to items 17 and 18, instances of a student altering her problem solving approach occurred only with items classified as having a Complex direction of problem solving structure. The following excerpt shows Student Q's initial response to item 17, including codes that indicate changes to the direction of problem solving in response to this item.

**Student Q:** *I'm going to-- I'm going to say choice two is not possible, because since it's acting at an angle, and since part two is curved, and it can only move in certain direction, the force isn't going to-- direction force where they're connected, isn't going to be the exact same as the force being applied on it, and there's going to have some sort of angle, so it wouldn't be directly up like this option has, so-- but there still would be a magnitude, since they aren't in line. So my choice, my answer is going to be whichever one says-- option one is possible, option two is impossible, which is B, [IR] and then, ... another possible choice, I would bounce between, would be E, can't say without more information, because we don't know how much force this is applying here, and we don't know if this isn't already compressed, or this spring at the top of part one is doing anything, or if it's just there, whether it's been stretched out already or if it's compressed, we don't know anything about that. So if there was another choice, I could bounce between, would be E, because we don't know how much force is at-- how much the force is being applied [NF],... but my option is going to-- my final choice is going to be B. that option one is possible, because the force is not in line with the contact force, the force that's being applied to part two is not in line with the contact force between part one and part two, which is the part in question. And so that would eliminate C and D, and then part two, I said would not be possible, because it wouldn't be at the same angle as applied force, since the contact part is at a different angle, so that would mean part A isn't-- or option A is false, so that's why I chose option B. The first option is possible, and the second option is impossible [IR].*

#### 4.3 Distractor Attractiveness

The Distractor Attractiveness theme includes three codes intending to describe the form of the problematic reasoning that leads to the selection of an incorrect response: (1)

Misconception, which indicates incorrect responses that align with common misconceptions, (2) Misapplication, which indicates incorrect responses that align with the misapplication of a scientific principle or when a principle is applied in an incorrect context, and (3) Partial, in which incorrect responses satisfy one aspect of a scientific principle, but do not satisfy the principle entirely. These codes were applied independently to the content analysis and the verbal data obtained through think-aloud interviews. Coding of the content analysis provides problematic reasoning that is expected based on the item design and possible problematic reasoning that would lead to the incorrect response. Coding of the verbal data obtained through think-aloud interviews provides problematic reasoning that is observed. In the following tables, distractor coding resulting from the content analysis is labeled as Expected Reasoning while distractor coding resulting from analysis of interview data is labeled as Observed Reasoning. The following findings are presented by CATS item groups and by item or item pairs when appropriate.

#### 4.3.1 Distractor Attractiveness of Less Difficult CATS Items

**Items 1 & 3:** Because these items are asking students to select a correct Free-Body Diagram (FBD) from options of alternate FBDs, the items focus more on students' knowledge of how forces should be represented in FBDs rather than the principles governing the manner in which forces would act on the given bodies. For this reason, these items were not coded using the thematic scheme. The correct responses for items 1 and 3, responses D and C respectively, are expected to indicate students' ability to correctly identify a free-body diagram representing the forces acting on a specific subsystem of a mechanism composed of multiple bodies and connections. Table 4.5

shows occurrences of incorrect responses by two students; the observations are shown in gray text because they don't adhere to the thematic coding scheme.

Table 4.5 Distractor Attractiveness Coding for CATS Items 1 & 3

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
1	A		(S) FBD that included an internal force and included weight in a tension force
	B		
	C		
	D		Correct Response
	E		
3	A		(O, S) FBD that included an internal force
	B		
	C		Correct Response
	D		
	E		

The following interview transcript excerpts show some evidence that students are trying to include an internal force in a FBD, but this may be due to an incorrect understanding of representational conventions rather than an error in physical reasoning. For both instances, the students are correct in their physical reasoning that the weight force acting on the mechanism would be balanced by a tension force, however because the tension force acts between members within the system and does not cross system boundaries, it would not be included in the FBD representation.

**Student S:** *"And similar to force TC, there would be a force TD acting upward, again, to keep the system intact. The TD acting upwards from Block 3... There would be a tension force in Cord E and F acting downwards, only because the gravitational force of Block 5 and 6 is not taken into consideration, because they are not part of the system... But in addition to the tension force, in Cord E and F would be a part of the weight of Block 4, just because it's not connected by a cord, but rather lays between two blocks."*

**Student O:** *"Looking at the diagrams, I know there's the weight down and the tension up for Weight 1 and TA, so that makes sense. I would rule out C and D"*

because they have the weight of 3 going down but they don't have the tension going up \_\_\_\_\_. Also, E has the Tension C going down instead of up. I think based on the Weight 1 and Tension A, *I think that the tension needs to be going up*. So that narrows my answers down to A and B, and the difference between A and B is the added weight of 4 and 5, and it doesn't show anything about 2, so that's kind of confusing why that is just left out, or 7.”

**Items 14 & 15:** These items ask students to select the direction of a reaction force for a pin-in-slot joint. The correct responses for these items, B and E, are expected to indicate students' ability to recognize that contact forces between a pin and slot always act in a normal or perpendicular direction. The distractors for these items are associated with two common errors: responses that align with the misconception that forces act along a member and the misapplication of equilibrium where forces act to balance an external force. Furthermore, each item included a distractor with an incorrect addition of a couple acting about the pin joint. These distractors were later identified through content analysis as potentially aligning with a misapplication of the principle of moments. The principle of moments roughly defines a moment as the product of the magnitude of a force acting perpendicularly to a pivot point and the distance between the force and the pivot point. A misapplication of this principle may occur when any force acting at a distance is used to reason the existence of a moment, even when a member is free to rotate about a pivot point. Table 4.6 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in six students' incorrect responses.

Table 4.6 Distractor Attractiveness Coding for CATS Items 14 &amp; 15

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
14	A	Misconception: Force along member	
	B		Correct Response
	C	Misapplication: Equilibrium	Misconception: Force in same direction as external force (L) Force in the same direction as external force
	D	Misconception: Force along member (slot) Misapplication: Moment	
	E		Misapplication: Moment (J, O, T) Force at a distance results in a moment
15	A	Misconception: Force along member Misapplication: Moment	Misconception: Moment (U, V) Moment as a contact force that prevents rotation of another body
	B		
	C	Misapplication: Equilibrium	
	D	Misconception: Force along member (slot)	
	E		Correct Response

Although not initially expected as a form of problematic reasoning associated with distractor C for Item 14, there is the possibility that this distractor may tap into a misconception that a force on a body must be in the same direction as external forces. The student responses to Item 14 show evidence of this as well as evidence of a misapplication of the moment principle, however the reasoning evident in students' responses to Item 15 do not align with a misapplication of moment. Instead, the responses seem to suggest that these students may have an underlying misconception of moment. While it is conceptually correct to think of a moment as a reaction that prevents rotation, the moment would act as a reaction about the pivot point of the body of interest. In this case, the students argued that since the pin was stopping the rotation of a contacting body about a different pivot point, a reaction moment would occur at the point

of contact. It is possible that the students reasoned that the nature of a contact force is dependent on the type of motion that is prevented through contact. Although not pressed further at the time of the interview, it would be useful to delve further into this reasoning as it seems that there is additional problematic reasoning that should be explored related to the conditions for reaction moments.

**Student L:** Because I'm just going by the fact that if you're applying the force by and since the slot is considered to be frictionless, once it moves downward, the force is angled at the same direction as the external force and since it's 10 degrees from the normal, it would be the same angle."

**Student J:** "it would have to be perpendicular because it can't be in the same direction because it would just move freely and there's nothing to provide a force, it's just open space. Oh the moment. Yeah, *I don't see any reason why this force wouldn't provide a moment about E because the only reason it wouldn't provide a moment if it is was pointed directly at A like into the point you're taking the moment about. So I think it would provide a positive moment.*"

**Student V:** "And then there'd be a reaction at A, which would act normal to the slot. So, I think because there can't be any forces in the-- on each side where the slot is empty, I guess, the horizontal length of it, so it would have to be... either answer B or E. And then, as far as reaction moments-- *I guess there would be a reaction moment at point A, as well because this has a tendency to fall down so it needs something that would counteract the action.* So, I would select B."

**Item 19:** Item 19 was designed to determine if students have mastery of the reaction forces at a pin connection and their representation. Students should be able to recognize that since a connection at a point is free to rotate, there can be no couple or reaction moment about the pin. Correct response A shows the correct representation of horizontal and vertical reaction forces that indicate one side of the force interactions preventing translational motion relative to the pin and the plate. Through content analysis, distractors were identified that may align with a misapplication of the principle

of moments. Students may incorrectly expect that forces drawn at a distance from the pin joint would result in a moment about the pin joint. Table 4.7 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in four students' incorrect responses.

Table 4.7 Distractor Attractiveness Coding for CATS Item 19

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
19	A		Correct Response
	B	Misapplication: Moment	Misapplication: Moment
	C	Misapplication: Moment	Misapplication: Equilibrium
		Misapplication: Equilibrium	(Q) Reaction force to balance applied force & Force at a distance results in a moment
	D	Misapplication: Moment	Misapplication: Moment (J) Force at a distance results in a moment Misconception: Moment (L) Rotation as evidence of moment
E	Misapplication: Equilibrium	Misapplication: Equilibrium (N) Reaction force to balance applied force	

Additionally for Item 19, content analysis revealed that students may incorrectly apply equilibrium and attempt to select reaction forces that would counteract the external force supplied by the rope keeping the plate static. The student responses for these items do show some evidence of reasoning that aligns with the misapplication label.

**Student Q:** Which is going to be a force coming 45 degrees from the pin. And I chose that one *because this plate is being held in place by an arm that's at 45 degrees*, and both of the forces are going to move it clockwise. So I assume that there would be the same force in the same direction as there would be here. The way it was-- the force was going to act. And so I chose-- actually, no, let's change that to C, *because there's going to be a moment since the force is applied directly at-- there's going to be a moment with this top force.*

**Student J:** And if you apply a force in this direction it would have an X and a Y component. *So I would say that there would be an external moment about that point depending on where the forces were applied.* But I don't think you can just



say that there wouldn't be any moment. I don't think that-- yeah, I would go with D because I think there would be a moment about the bottom point.

**Student L:** So, I took option D because it's a pin joint so basically it applies the force on the X and the Y *and since it's a pin then you can rotate it off that so there could be moment*

**Student N:** Here, we're going to have, on the first pin, with the rope holding it will have a force acting there, holding the block in place and at the pin that's directly connected to the block will have just the two basic X and Y forces ... so for B, C, and D, we have a moment there which there's no opposing forces to the block's movement so those would not be included. *It would be 45 degrees because it's the same angle as the rope holding it in place.*

#### 4.3.2 Distractor Attractiveness of More Difficult CATS Items

**Items 4 & 5:** These items were designed to target students understanding of forces at a pin when separated for analysis. In this context, students should recognize that the forces between the bodies should be equal and opposite. D is the correct response for both items and in both cases, the diagrams for this response shows forces of equal magnitudes and in opposite directions. The distractors for this item were designed to align with an overgeneralization of two-force members in which forces may only act along bodies. Table 4.8 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in ten students' incorrect responses.

Table 4.8 Distractor Attractiveness Coding for CATS Items 4 &amp; 5

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
4	A	Misapplication: Force perpendicular to member	Misapplication: Equilibrium (O, Y, AA) Resultant forces opposite of applied forces in problem stem Misapplication: Equilibrium (L, N, Q) Resultant forces opposite of applied forces in problem stem Correct Response
	B		
	C		
	D		
	E	Misapplication: Force along (parallel to) member	
5		Misapplication: Force along (parallel to) member	Misapplication: Force along (parallel to) member (Q) Force along the frame to act on connection Misapplication: Equilibrium (N, O, Q) Resultant forces opposite of applied forces in problem stem Correct Response
	A		
	B		
	C	Misapplication: Force perpendicular to member	
	E		

In the interview data, one student demonstrated some evidence of the expected problematic reasoning, however multiple students showed evidence of reasoning that may be categorized as a misapplication of equilibrium. In the student responses coded as misapplications of equilibrium, students select distractors wherein the forces shown in the incorrect responses appear to balance the forces shown in the problem stem diagram.

**Student AA:** “I was looking at based on *if you sum the forces acting on that bar, what the final vector in a sense would be*. And that doesn't seem to be correct... In that case, there have to be vertical forces on it, since even if the vertical forces on the bottom bar cancel out, there's no way the vertical forces on the top bar cancel out. So that means it can't be D... *because if you're looking at the reaction, it would end up being B, because the opposite of this—the resulting of the forces on the bar, the opposite is the reaction against it by the pin, so that would be that.*”

**Student Q:** “I finally chose option C, because we have the beam, the straight beam. *I said it would act opposite of what the load is*, so I put everything in equilibrium and said that there would be a greater load, there would be greater force in this direction, or I guess I want to say greater force going down, because of the downward load would be greater than the load pointing up, with the second load that's acting at an angle in the Y direction, that's going to be lesser, so that's why I chose the force at-- the force in option C as the correct option”

**Student O:** *"I would guess that the force would be going opposite of the two black arrows shown. So to the right, because I would think that would kind of balance the system out. But then the bar itself is angled, so maybe instead of up, that would just be angled to the right...because the forces are just like out in the air, and to me, they need to be like going along that frame, because otherwise, how are they pushing on the pin?"*

**Items 7 & 8:** Students are asked to select a correct equivalent system from options with a variety of forces and couples. When these items were designed, the distractors were meant to target a common tendency for students to only partially apply equilibrium conditions to equivalent systems. The correct responses to items 7 and 8, responses E and B respectively, are the only response systems that would maintain equilibrium as determined through application of force and moment balance considerations. All of the distractors can be classified within the Partial category as they each fulfill one part of the equilibrium conditions: equivalent force or equivalent moment, but not both. Table 4.9 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in seven students' incorrect responses.

Table 4.9 Distractor Attractiveness Coding for CATS Items 7 &amp; 8

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
7	A	Partial: Force equivalent, not moment	
	B	Partial: Moment equivalent, not force	Partial: Moment equivalent, not force (P, R, S, T) Equivalent moment evaluated from point of couple action
	C	Partial: Moment equivalent, not force	
	D	Partial: Force equivalent, not moment	
	E		Correct Response
8	A	Partial: Moment equivalent, not force	
	B		Correct Response
	C	Partial: Moment equivalent, not force	Partial: Moment equivalent, not force (P, R, S) Equivalent moment evaluated from top point
	D	Partial: Moment equivalent, not force	
	E	Partial: Force equivalent, not moment	

Interestingly all of the students who incorrectly answered these items chose the same distractor. All of the instances in which students only partially applied equilibrium conditions consisted of reasoning in which students applied moment balance considerations without ensuring equivalent force conditions. Additionally, there seems to be some evidence that a “simpler” equivalent system is more attractive; student mentioned simpler load conditions as an argument for why one system is more correct than a seemingly equivalent system. However, further research would be needed to explore problematic reasoning related to simplicity of load conditions and whether or not this concept of simplicity is unique to items on equivalent systems.

**Student S:** “So now what we could do is calculate the moment about the second point from the left on the plane...Going back to option B, *calculating the moment at the second point from the left would be two Newtons times the perpendicular distance, which is 100 millimeters, which would be 200 Newtons*

*per millimeter*. But the difference between this option and the option in the question is that the moment at option B would act-- no, I'm sorry. There is no difference between the option B and the diagram in the question.

**Student R:** "Well e or c-- it's only one load; so it's more like simple to show...*If you sum the moments about this point*, it would be a moment of 400 Newton meters. So technically it would be the same reaction.

**Items 22 & 23:** These items are intended to access students' understanding of static friction within the context of sliding blocks. The correct responses to items 22 and 23, B and D respectively, are expected to indicate students' ability to recognize static friction conditions and correctly calculate a friction force below the maximum static friction value that would maintain equilibrium for a system. Distractors were designed to align with the common error of a misapplication of the static friction formula wherein the friction force must equal the maximum static friction value ( $f = \mu N$ ) even when equilibrium is maintained at a lesser value. Other distractors were designed with variations of this error in which the maximum friction force is added to or subtracted from applied force values. Table 4.10 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in 11 students' incorrect responses.

Table 4.10 Distractor Attractiveness Coding for CATS Items 22 &amp; 23

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
22	A		
	B		Correct Response
	C		(S) Vector sum of applied force and max static friction force
	D		
	E	Misapplication: Friction	Misapplication: Friction (K, M, V, W) Friction force must equal $\mu N$
23	A		
	B	Misapplication: Friction	
	C		
	D		Correct Response
	E	Misapplication: Friction	Misapplication: Friction (P, R, S, U, V, W) Friction force must equal $\mu N$

Nearly all of the errors observed in the student interviews aligned with this expected misapplication of the friction force formula and students provided evidence of reasoning that included a calculation of the maximum static friction force without considering the forces needed to maintain equilibrium.

**Student M:** we know that the coefficient of friction is *.5 assuming that the  $F$  is equal to  $\mu$  times the normal force. And  $.5$  times  $20$ , which is  $10$ , so the friction force is  $10$  Newtons acting to the right, and that is answer E.*

**Student W:** And actually, I think that's all I need because I know that the force of friction equals just  $\mu$  times  $N$ ,  $N$  being the normal force, and I know that the normal force has to be  $90$ . We're just looking for the friction force of the ground, correct? Yeah, so just by the floor on the lower block. *So I would just say that's  $0.2$  times  $90$ , which is  $18$ , I think. Yes,  $18$ .*

#### 4.3.3 Distractor Attractiveness of Items with Difficult Cognitive Attributes

**Item 17:** This item was designed to determine if students understand that all contact forces are perpendicular or normal to the tangential surface or point of contact. The correct response, D is expected to indicate students' ability to rule out both presented result situations as they do not include the correct solution of a single force perpendicular

to the point of contact. Upon review of the item in content analysis, the distractors may align with some common misconceptions that could attract students beyond the issue of contact force directionality. Content analysis revealed that students may select distractors with a vertical contact force if they hold the misconception that the reaction/contact force would need to be in the same direction as the applied force. Additionally, student may select distractors with a couple at the contact point if they misapply the principle of moments and assume that a force at a distance implies the presence of a moment. Table 4.11 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in six students' incorrect responses.

Table 4.11 Distractor Attractiveness Coding for CATS Item 17

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
17	A	Misapplication: Moment Misconception: Force in same direction as external force	
	B	Misapplication: Moment Misconception: Force in same direction as external force	Misconception: Moment (J, Y) Rotation as evidence of moment Misconception: Force in same direction as external force (O) Force in same direction as external force
	C		Misapplication: Equilibrium (L, N) Upward reaction to balance downward weight force (X) Vertical force as component of normal force
	D	Correct Response	
	E	Correct Response	

The examples of student reasoning as collected from the interviews did not align for the most part with the expected errors as identified in the content analysis. In one case a force misconception was observed, however students' reasoning also contained unexpected misconceptions: rotation as evidence of a moment (rotational version of the

impetus theory) and an unexpected misapplication of equilibrium in which the upward contact force was described as balancing the downward weight force.

**Student Y:** ...because there's a force acting on Part B which will make the Part B rotate counterclockwise and that would make Part A to rotate also. *And so there will be like a motion, circular motion at the contact of two points, so there will be a moment there.*

**Student O:** In Answer 2, it just shows the force arrow moved over a little bit to the point of contact between Part 1 and Part 2. I think that's possible *because the force is in the Y direction and it is touching Bar 1 at that point.*

**Student N:** Yeah I think this force is part one I guess; *the only force that's really acting at that point is MG (weight force calculation). So to oppose that force, all we have-- all you need is that one force going up so I'm going to say two is possible, one is impossible, C.*

**Items 18:** Item 18 was designed to determine if students have mastery of the reaction forces at a pin joint and their representation. Students should be able to recognize that since a point joint is free to rotate, there can be no couple about the pin. The correct response, C is expected to indicate students' ability to rule out a response that includes a couple or reaction moment about the pin, but include a reaction force that can be any vector sum of horizontal and vertical reaction forces. Through content analysis, distractors were identified that may align with a misapplication of the principle of moments. For both items, students may incorrectly expect that forces draw at a distance from the pin joint would result in a moment about the pin joint. Additionally, there were two instances of students' problematic reasoning that did not conform to the coding structure. Table 4.12 includes the coded expected and observed reasoning that would lead to selecting specific distractors, as well as the reasoning identified in three students' incorrect responses.



Table 4.12 Distractor Attractiveness Coding for CATS Item 18

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
18	A	Misapplication: Moment	Misapplication: Moment (L, Q): Force at a distance results in a moment
	B	Misapplication: Moment	
	C		Correct Response
	D		
	E		(O) Assumes that values are needed to apply equilibrium conditions

**Item 27:** Students are asked which additional load would keep an object in equilibrium. The correct response for this item, A, is expected to indicate students' ability to recognize why the diagram presented in the problem stem is not in equilibrium, and to select an additional load that would balance the forces and moments acting on the body. Response A is a single vertical force that could balance a vertical component of the force applied at point P, assuming that the horizontal component of the force applied at point P is balanced by the force applied at point Q. Due to the colinearity of the opposing forces, this arrangement of loads would not result in a net moment, and would lead to equilibrium conditions. Students would need to recognize that the distractor D, although a similar vertical force, would be located off-set from the balancing vertical component of the force applied at point P and would result in a net moment, not resulting in an equilibrium state. Similarly to Items 7 & 8, the distractors for this item were designed with the idea that students would either correctly apply both aspects of the equilibrium equations (force and moment) or incorrectly apply one without the other. In this manner each of the distractors for this item were coded as partial application of the equilibrium equations. Table 4.13 includes the coded expected and observed reasoning

that would lead to selecting specific distractors, as well as the reasoning identified in five students' incorrect responses.

Table 4.13 Distractor Attractiveness Coding for CATS Item 27

CATS Item	Distractor	Expected Reasoning	Observed Reasoning
27	A		Correct Response
	B	Partial: Forces not balanced Partial: Forces not balanced	Partial: Forces not balanced (Q) Forces not balanced, loads balanced on points
	C		(O, X, Y) Moment selected to balance applied forces
	D	Partial: Forces balanced, not moment	Partial: Forces balanced, not moment (N) Forces balanced, not moment
	E	Partial: Forces not balanced	(J) Selects response to balance assumed moment

In some cases, the student interviews align with this expectation. However, there were additional errors that emerged from the analysis of student interviews. Students also provided evidence of reasoning that suggests problematic reasoning relating rotational and translational forces. In three instances, students selected distractors that included couples with some explanation that the moments would balance the applied forces. Further research is necessary to tease out why students are confounding the effects of translational and rotational forces.

**Student Q:** I think any of them could put it into equilibrium, but because Pin R isn't on the same plane as P and Q, then my reasoning behind this is if you included a force or a moment at either P or Q, then that would put the other one, either P or Q, the opposite one in equilibrium. But I don't think R would be in equilibrium with the rest of them. *So I think you have to put an additional load at Point R to put the whole entire thing in equilibrium.* So there's only one choice that has anything acting upon R, so that's Choice C.

**Student O:** You would need a positive X force and a positive X and Y force with the same angle and magnitude as P, which would be kind of two different

vectors. You would need this and this. *So since there's two, that would make me think that maybe you'd need C or E, since they're moving in two different directions.* They're actually moving in other directions. So I'm just going to go with a final answer of C

**Student N:** I was also considering A because I was thinking that the X component of the P was taken care of and that the only thing left we had to worry about was the Y, but since both the Y and the X are acting at P, *you really need a Y on the Q in order to balance it out.*

**Student J:** So I guess you want the moment about any point to be zero. So if you applied it at P the moment about Q wouldn't be zero ... my gut feeling would be to go with applying that moment at Q (E) *just because I guess they would cancel each other out because one's positive and one's negative.*

#### 4.3.4 Summary of Distractor Reasoning

In summary of the distractor attractiveness codes applied to the items in the previous sections, Table 4.14 includes a listing of the detailed forms of problematic reasoning aligning with each distractor code, as well as the items containing the distractor codes.

Table 4.14 Summary of Expected and Observed Problematic Reasoning Coded to Item

## Distractors

Distractor Code	Problematic Reasoning	Items – Observed and (Expected)		
		Less Difficult	More Difficult	Difficult CAs
Misconception	<b>Force:</b> Force in the same direction as an external force	14		17
	<b>Force:</b> Force along member	(14/15)		
	<b>Moment:</b> Rotation as evidence of moment	19		17
	<b>Moment:</b> Moment as a contact force that prevents rotation of contacting body	15		
Misapplication	<b>2 Force Member:</b> Force parallel to member; Force perpendicular to member		(4), 5	
	<b>Moment:</b> Force acting at a distance results in a moment	14, 19		(17), 18
	<b>Equilibrium:</b> Reaction force to balance applied force; upward reaction force to balance downward weight force	(14/15), 19		17
	<b>Equilibrium:</b> Resultant forces opposite of applied forces in problem stem		4/5	
	<b>Friction:</b> Friction force must equal $\mu N$		22/23	
Partial	<b>Equilibrium:</b> Moment equivalent, not force		7/8	
	<b>Equilibrium:</b> Force equivalent, not moment		(7/8)	
	<b>Equilibrium:</b> Forces not balanced			27
	<b>Equilibrium:</b> Forces balanced, not moment			27

As seen in Table 4.14, less difficult items include distractors that code to misconceptions and misapplications. Although these items tend to focus on reactions at a pin or contact point, students often select incorrect responses that include reaction moments. More difficult items include distractors that code to partial applications equilibrium, either applying force balances independent of moment considerations or applying moment balances independent of force considerations, among other forms of problematic reasoning.

It is also interesting to note that Item 17 is one of the most difficult items by psychometric measure; however the distractors coded to forms of problematic reasoning are more similar to characteristics of less difficult items. Item 18, a conceptual item pair

with 17, has a much lower item difficulty and also codes to similar forms of problematic reasoning common to less difficult items.

#### 4.3.5 Distractor Attractiveness by Item Group

By consolidating the individual item codes, a comparison may be made about how distractor attractiveness as a theme across the different item groups. Table 4.15 presents a summary of the number of incorrect responses observed by item distractor along with the distractor attractiveness codes indicated expected problematic reasoning from the content analysis and observed reasoning from think-aloud interviews.

Table 4.15 Distractor Attractiveness by Item Group, Codes are Indicated as Misconception (C), Misapplication (A), and Partial (P)

CATS Item Groups	Distractor Coding																		
Less Difficult Items	1		3		14		15		19										
Student Response Tally	1		2		1	3	2		1	3	1								
Content Analysis Interviews					C	A	C	A	C	A	A	C	A	A/C	A/C	C			
More Difficult Items	4		5		7		8		22		23								
Student Response Tally		3	3	3		4		3		1	4		6						
Content Analysis Interviews	A			A	A	A		P	P	P	P	P	P	P	P	P	A	A	A
Items with Difficult CAs			17			18					27								
Student Response Tally		2	4		2		1		3		1	1							
Content Analysis Interviews	A/C	A	C		A	A		P	P		P	P							

Comparing the expected and observed forms of problematic reasoning leading aligned with distractor selection may provide some evidence supporting or countering psychometric models that include errors associated with specific CATS items. It is evident that in some cases the observed codes conform reasonably well with the expected codes, such as with items 7, 8, 22, 23, 18, and 27. However, evidence of students' problematic reasoning and distractor selection from items 14, 15, 19, 4, 5, and 17 does not align well with expected reasoning. Although not easily seen in Table 4.15, it is also important to note the instances when students' observed problematic reasoning was not captured by the distractor attractiveness coding scheme. Items 17, 18, 22, and 27 all included instances of observed student responses wherein the problematic reasoning expressed did not conform with the coding scheme for distractor attractiveness

It is interesting to note that types of distractors codes as Misconceptions or Misapplications of occur across items with a range of item difficulty. However, the Partial code only appears for more difficult items and items with difficult cognitive attributes.

#### 4.3.6 Distractor Attractiveness by Student

Examining forms of problematic reasoning leading to incorrect responses by student may provide some description of the nature of student errors across a test. Table 4.16 below displays the incorrect responses for each student using the distractor attractiveness thematic codes. Because each student did not respond to each CATS item, the items that were not included in the booklet presented to the student are grayed out.

Table 4.16 Distractor Attractiveness by Student, Codes are indicated as Misconception (C), Misapplication (A), Partial (P), and Other (O)

Student Group	Less Difficult Items				More Difficult Items				Items with Difficult Cognitive Attributes					
Item	1	3	14	15	19	4	5	7	8	22	23	17	18	27
<b>Booklet 1</b>	K									A				
	M									A				
	P							P	P		A			
	R							P	P		A			
	S	O	O					P	P	O	A			
	T			A				P						
	U				C						A			
	V				C						A	A		
	W										A	A		
<b>Booklet 2</b>	J			A		A						C		A/P
	L			C		C	A					A	A	
	N					C	A	A				A		P
	O		O	A		O	A	A				C	O	O
	Q					A	A	A					C	P
	X											O		O
	Y						A					C		O
	Z													
	AA						A							

As one may expect, there are generally less incorrect responses for the less difficult items. Additionally, in some cases the same student makes a similar type of error across similarly designed items, such as with items 4 & 5 and 7 & 8. However, in other cases the same student does not make the same error across similarly designed items. These inconsistencies in form of problematic reasoning leading to incorrect responses occur with both less difficult items and more difficult items.

Instances in which students' problematic reasoning does not conform with the distractor attractiveness coding scheme is more evident in Table 4.16 and indicated with

an O. While the forms of problematic reasoning that students apply in incorrect responses with respect to the coding scheme are inconsistent, there does appear to be some level of consistency in the other responses. Looking at just the other, problematic reasoning that does not conform to the coding schemes seems to be exhibited by a few specific students, specifically students S, O, and X. Students S and O may be characterized as low-performing students while student X may be described as a high-performing test taker. Therefore, it may be suggested that students' observed problematic reasoning not captured by the coding scheme is independent of the students' overall test performance.

#### 4.4 Summary

The findings presented in this chapter described in detail evidence from a content analysis of CATS items and think-aloud interviews as a means of providing context for students' reasoning in response to CATS items within groups varying by difficulty. The themes that emerged from thematic analysis of the collected data are (1) Direction of Problem Solving and (2) Distractor Attractiveness.

In summary of the findings from the direction of problem solving theme, two levels of coding were established and applied. Firstly, applying item-level codes showed that less difficult items are more likely to be characterized as Forward problems in which a single correct response can be determined from the provided problem stem and more difficult items more likely to be characterized as Reverse problems that do not pose a problem stem with a single correct response and require evaluation of all responses. Secondly, applying student response-level codes showed that students are nearly twice as likely to demonstrate a change in problem solving approach with Reverse items.



Additionally, items 17 and 18 were coded as Complex, meaning that the problems included two sets of responses that required evaluating multiple conditions of responses. For these items, students demonstrated a change in problem solving approach, but with differing frequency. One student was observed changing direction of problem solving twice in response to both Complex items.

In summary of the findings from the distractor attractiveness theme, the expected errors associated with each item distractor and the problematic reasoning that students demonstrated in response to each distractor were characterized by the application of three codes: misconception, misapplication, and partial. It is interesting to note that items included in the less difficult group do not have distractors that were designed to tap into student errors that exhibit a partial application of a scientific principle. Distractors that align with misconceptions were only found with less difficult items and item 17. In some cases, the observed reasoning that led to distractor selection aligned well with expected problematic reasoning, specifically those items with distractors characterized as partial applications of scientific principles. Additionally, students' problematic reasoning as described by the distractor attractiveness coding scheme is fairly consistent across some item pairs, specifically items 4 & 5 and 7 & 8. In the remaining cases, student reasoning is inconsistent across even related items, regardless of difficulty.

## CHAPTER 5. CONCLUSION

### 5.1 Introduction

The purpose of this dissertation study was to build upon an aspect of assessment design that makes meaning of test scores. In doing so, an exploratory qualitative study was used to elaborate upon implications and limitations of existing statistical models associated with CATS. Prior research had determined psychometric measures that characterize student performance on CATS based on multiple theoretical measurement models including classical test theory and item response theory. Common to these approaches is the determination of item difficulty, a parameter that is used to distinguish which items are more difficult than others. The present study was intended to provide rich description of what makes some CATS items more difficult than others within the context of statics and based on students' reasoning in response to CATS items.

Specifically, the research question guiding this study was:

- How does student reasoning in response to CATS items explain variance in item difficulty across test items?

To answer this question, think-aloud interviews were conducted in combination with a content analysis of selected CATS items. Thematic analysis was performed on interview transcripts and CATS development and evaluation documentation. Two themes emerged as possible explanations for why some CATS items are more difficult

than others: (1) a Direction of Problem Solving theme describes the direction of reasoning required or used to respond to CATS items, and may also provide some description of students' reasoning in response to determinant and indeterminant items; and (2) a Distractor Attractiveness theme describes problematic reasoning that is targeted and observed in incorrect CATS responses.

The following chapter responds to the dissertation research question through discussion of the findings from the thematic analysis, conclusions based on findings, a proposed theory of difficulty for CATS and how this relates to Cognitive Load Theory, commentary on the intersection of item design, student reasoning and item difficulty and how this relates to the Evidence-Centered Design Framework. Finally, implications and recommendations are provided including commentary on CATS as an instructional tool and concept inventory development.

## 5.2 Discussion of Findings from Thematic Analysis

Overall, the findings from the direction of problem solving theme suggest that item difficulty may be contributed to by the format and design of the items included in the concept inventory. More difficult items are associated with a Reverse item design, while less difficult items are associated with a Forward item design. Furthermore, students are more likely to change problem solving approaches with more difficult items. This may suggest that the ability to recognize how to go about correctly solving CATS items may be contributing to item difficulty.

To summarize the findings from the distractor attractiveness theme, more difficult items are more likely to include items designed with distractors that tap into errors resulting from a partial application of equilibrium conditions, while distractors that align

with misconceptions and misapplications of scientific principles are found in items of varying difficulty. Students' incorrect responses to the set of items included in the more difficult items group were more likely to be focused on one or two strong distractors that aligned with expected reasoning for the items. These items were also found to have the most consistency in terms of students' problematic reasoning across items of the same target concept. Although there was more variety found in the distractor designs and how students responded to the items in the less difficult and items with difficult cognitive attributes groups, evidence is presented in section 5.2.2 suggesting that the specific problematic reasoning targeted by concept inventory items may contribute to item difficulty.

#### 5.2.1 Discussion of Findings from Direction of Problem Solving Theme

In comparing the item-level design of the selected CATS items with respect to how respondents would likely progress in the direction of problem solving, it was found that less difficult items more likely to be characterized as Forward items, while more difficult items more likely to be characterized as Reverse items. This implies that CATS items with a problem stem leading to a single, correct response may be less difficult than items that require evaluation of all responses. I contend that this is due, at least in part, to an increase in complexity and higher level thinking required of the format of Reverse items. The Forward or determinant items may be characterized as testing the ability to recall knowledge whereas the Reverse items require students to evaluate and consider the specific conditions that apply to the problem. In the context of Bloom's taxonomy as a basic framework for types of thinking, evaluation is of a higher order than recall in how learners use knowledge (Marzano, 2001). The Reverse items of CATS may elicit a

higher level or more complex form of reasoning than the Forward items, contributing to item difficulty.

In comparing coded student responses across items, students are nearly twice as likely to demonstrate a change in problem solving approach with Reverse items. Additionally, for items 17 and 18, students demonstrated a change in problem solving approach with both Complex coded items, but with differing frequency. One student was observed changing direction of problem solving twice in response to the Complex-coded items. In combination, these findings may suggest that item difficulty is related to students' ability or inability to recognize the best directional problem solving approach when posed with CATS items. In conversation with Paul Steif, the idea arose that items 4 & 5 in particular "look" like traditional statics problems. Engineers are taught to approach problems in a uniform and linear manner, progressing from stages of identifying the problem statement, listing unknown variables, selecting appropriate principles and fundamental equations, etc. (Litzinger et al., 2010). Due to this, it may be possible that concept inventory items that contain problem stems without a single correct answer and require evaluation among responses are more difficult than determinant items, especially for those items that resemble traditional engineering problems.

### 5.2.2 Discussion of Findings from Distractor Attractiveness Theme

In comparing among the less difficult items, this group included items that focus on reactions at pin joints and contact points. The distractors for these items may be described as variations on reaction conditions that may align with specific common errors, misconceptions, or misapplications. Using the coding scheme of the distractor

attractiveness theme, all of the less difficult items had distractors that coded to misconceptions and misapplications.

Distractors that were coded as misconceptions were only found with less difficult items, with the exception of item 17. It appears that distractors that target the specific misconceptions identified in this study may be associated with less difficult items. In examining the problematic reasoning that was identified in student responses, there may be some connections to previously identified misconceptions in physics and engineering mechanics:

- *Reaction force in the same direction as an external force*: this identified instance of problematic reasoning may align with previous studies that have found students' tendency to view force as something that transfers from one body to another. However there are multiple views as to how this problematic reasoning is classified. It has been described as a property of an impetus theory (Clement, 1982), a "force as mover" p-prim (diSessa, 1993) or due to a substance-based view of force (Reiner et al., 2000). This misconception may also be reinforced by common illustrations of force in a single dimension.
- *Rotation as evidence of a moment & Moment as a contact force that prevents rotation of another body*: Both of these examples of problematic reasoning seem to suggest some underlying misconceptions of moment. Previous studies have noted that students have difficulty with several aspects of rotational motion, including relationships between force imbalance and rotation, and moment of inertia (Montfort, Brown, & Findley, 2007; Ortiz et al., 2005; Rimoldini & Singh, 2005). Additionally, these misconceptions also map to the existing statics

common errors 7 and 8 as defined by Steif and Dantzler (2005). While this study may serve as an introduction to some examples of problematic reasoning related to misconceptions of moment as applied to mechanisms under static equilibrium, there appears to be a substantial gap in knowledge regarding misconceptions of couples and moment that should be addressed.

The observed problematic reasoning for the less difficult items did not always match the expected reasoning for each distractor. This may suggest that additional considerations should be taken into account when using student responses to gauge problematic reasoning as part of instruction. However, this may also be due to limitations in the thematic coding scheme that does not capture nuances in student reasoning around misconceptions. For example, it is difficult to say in the present study if a student incorrectly responds to item 14 with a reaction force in the same direction as the applied force, whether this problematic reasoning is due to an underlying assumption that force is carried by objects and is passed from one body to another, or if the student is incorrectly attempting to balance the applied force through a misapplication of equilibrium. Because these nuances in reasoning are more difficult to tease out in the categories of misconceptions and misapplications, the evidence of distractor attractiveness for the less difficulty items is less convincing than for the more difficult items.

In comparing among the more difficult items, most of the distractors were coded to a partial application of equilibrium in which either force or moment balances are taken into account without reference to the other. The more difficult items also contained

distractors with higher occurrences of consistent reasoning and single highly-attractive distractors. In other words, the student responses to the more difficult items were more likely to be on the same one or two distractors and for the same reason. This potential relationship between more difficult items and highly attractive distractors should be confirmed with a larger data set.

In some cases, the observed reasoning that led to distractor selection aligned well with expected problematic reasoning, specifically those items with distractors characterized as partial applications of scientific principles. As previously discussed, the tendency for students to apply principles of equilibrium incompletely is well documented in statics learning literature (Litzinger et al., 2010; Newcomer & Steif, 2008a; Ortiz et al., 2005; Steif & Dantzler, 2005); the findings from this study support this observation. In the context of CATS, the distractors coded to the partial application of a scientific principle occurred when either force or moment were balanced without reference or consideration of the other. Statics is generally the first course in which students need to consider both force and moment balances as conditions for equilibrium. Previous research suggests there is a significant difficulty inherent in needing to apply both force and moment balances and that statics students have a tendency to apply one without the other (Litzinger et al., 2010; Newcomer & Steif, 2008a). Despite the limited strength of the distractor attractiveness findings for less difficult items, the strength of the observed association between more difficult items and distractors that align with partial applications of scientific principles suggests that the distractor form and specific problematic reasoning targeted by a CATS item may contribute to item difficulty.



When looking at individual students' responses as shown in Table 4-16, students' problematic reasoning is fairly consistent across some item pairs, specifically items 4 & 5 and 7 & 8. In the remaining cases, student reasoning is inconsistent across items, regardless of targeted concept or difficulty. Students' inconsistent problematic reasoning even across related items is not as surprisingly as it may initially appear. A noted study previously showed that novices are more likely to focus on contextual features of problems rather than underlying principles (Chi, Feltovich, & Glaser, 1981) and may not recognize item pairs as related. This finding also provides some agreement with a previous study that found inconsistencies in students' problematic reasoning across similar forms and contexts of statics problems (Newcomer, 2010). This finding may suggest that students are largely inconsistent in the form of problematic reasoning applied to seemingly similar items by concept.

### 5.3 Conclusions

CATS as a statics concept inventory has performed well in psychometric evaluations with regard to concept structure and other forms of validity evidence showing that the test measures what it intends to. Based on these previous studies, an instructor may be fairly certain that a student's correct response to a CATS item would indicate the student's mastery of a correlating concept. However, the previous studies did not provide as useful a set of inferences based on incorrect responses.

The first conclusion from this study's findings is primarily drawn from the Direction of Problem Solving Theme. In addition to the targeted concept for CATS items, item design and format contribute to item difficulty. For example, students who incorrectly respond to CATS items grouped on Newton's 3rd Law (4, 5, 6) may not

necessarily have difficulty with the concept of those items, but specific item features that contribute to the difficulty of the items.

Also pertaining to inferences based on incorrect responses, the second conclusion from this study is primarily drawn from the Distractor Attractiveness Theme. The types of distractors and forms of problematic reasoning that align with items and distractors also contribute to item difficulty. For example, one may assume that with a collection of statics concept items similar to those of CATs, items that include distractors resulting from a partial application of equilibrium would be difficult for a similar test-taking population.

Finally, findings from the Distractor Attractiveness Theme also produced the third conclusion. Patterns of student errors are not necessarily consistent by concept, as there may be item design features that influence student reasoning. For example, the problematic reasoning that led a student to incorrectly select a specific distractor for item 14 is not necessarily the same problematic reasoning that would lead the same student to an incorrect response for item 15. Although both items are grouped as Pin-in-slot items, students may not recognize the similarity of these items and apply differing problematic reasoning in incorrect responses.

### 5.3.1 Theory of Difficulty for CATS

Based on the findings from this study, a theory of difficulty for CATS can be described as an issue of compounding complexity, specifically the complexity inherent in the domain content compounded with the complexity of the item design. With regard to the complexity of context or item design, the findings suggest that indeterminate items may be more difficult than determinant items. As described in first conclusion, the

format of specific CATS items were associated with differences in item difficulty. Additional complexity of higher order thinking may contribute to the difficulty of items based on the direction of problem solving required to respond to the item. Extending that logic, the context of item problem stems and distractors that are more complex may also contribute to difficulty. For example, items that include multiple bodies and connections may be more difficult than items that require analysis of a single body.

With regard to the complexity of content, statics may be viewed as more difficult than introductory physics because of added complexity. The increased complexity of statics can be accounted for by the need to consider both force and moment balances (Litzinger et al., 2010; Newcomer & Steif, 2008a; Ortiz et al., 2005), as well as the need to examine connections between and among bodies within mechanisms (Steif & Dantzer, 2005). The content of individual items that is more complex may contribute to difficulty. As described in the second conclusion, items in which students have the opportunity to select distractors that only partially apply a scientific principle are more difficult. For example, items in which students must consider both force and moment balances may be more difficult than items that require analysis of force and moment independently.

### 5.3.2 Relation to Cognitive Load Theory

In addition to aligning with a common sense idea that complexity contributes to difficulty, the proposed theory of difficulty for CATS also aligns with theories of cognitive load. Cognitive load theory as referred to in this context describes how the load imposed on working memory (Sweller et al., 2011) or cognitive capacity is finite.

Just as the proposed theory of difficulty for CATS includes the compounding effects of content complexity and context complexity, cognitive load theory differentiates

between intrinsic cognitive load and extraneous cognitive load. Intrinsic cognitive load relates to the difficulty of the content: complexity and interactivity: material with multiple elements and high levels of interactivity requires use of more of the available cognitive capacity and is more difficult than simpler material (de Jong, 2010).

Extraneous cognitive load relates to the manner in which information is presented or displayed. Multiple forms of information formats and the extent to which multiple pieces and forms of information need to be integrated consume cognitive resources at higher levels. Sources of intrinsic and extraneous cognitive load have a compounding effect, just as the sources of item difficulty seem to have a compounding effect.

In the context of CATS, it is clear that items which require students to reason force equilibrium and moment equilibrium simultaneously are more difficult than items that are focused solely on force reactions. Adding to this, student errors among the less difficult items often arose when students' incorrectly presumed the existence of a couple. In effect, these students may have been making those items more difficult by trying to apply both equilibrium conditions when the context only required a balance of forces.

For most of the findings from this study, it appears as though an increase in complexity or cognitive load may account for potential sources of item difficulty. However, there is one example of disconfirming evidence that should be addressed. The items that were coded as having a Complex item format, items 17 and 18 have very different item difficulty values, but did not perform differently with respect to the thematic analysis. If the added complexity of the Complex item design is independent of the targeted item concept and the reasoning evident in student responses to those items, they should have similar values of item difficulty. However, it is possible that the

differences between items 17 and 18 are not observable with analytic approach of this study and require further research to understand why item 17 is much more difficult than item 18. Some test design research has argued that the added cognitive load of Complex multiple-choice items should prohibit their inclusion in assessment design. However, there is the possibility that population characteristics may mitigate this effect. Similarly to medical education literature (Case & Swanson, 2001), complex items may not pose an excessive cognitive load for engineering students.

### 5.3.3 Interrelationship among Item Design, Student Reasoning and Difficulty

As described in previous section, the proposed CATS theory of difficulty hinges on the idea that both the content of the concepts that are tested and the context of items within a test contribute to difficulty of concept inventory items. From this it is clear that there is an interconnected relationship among test item design, student reasoning and item difficulty. This interrelationship may be seen as a projection of the assessment triangle in which item design correlates with observation, student reasoning aligns with cognition, and item difficulty is a single construct within interpretation (Figure 5.1).

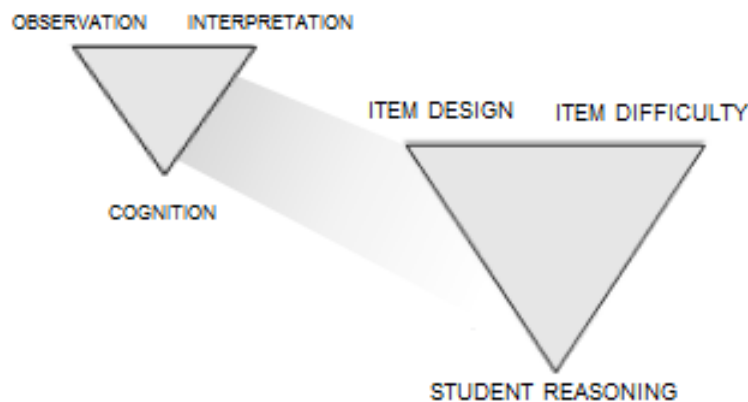


Figure 5.1. Interrelationship of Item Design, Student Reasoning, and Item Difficulty

The way in which item design and student reasoning were examined as potential sources of item difficulty holds potential for manipulation of item design as a means of targeting specific aspects of student reasoning and difficulty. In this way, the findings from this study may also be understood as establishing a framework that can be used to design additional statics concept items that function similarly to CATS.

#### 5.3.4 Relation to Evidence-Centered Design Framework

Without explicitly applying the Evidence-Centered Design Framework in the exploratory analysis, the findings from this study align with aspects of the ECD Framework, specifically the contributing elements of the Conceptual Assessment Framework (Figure 5.2) and the interrelated nature of the elements.

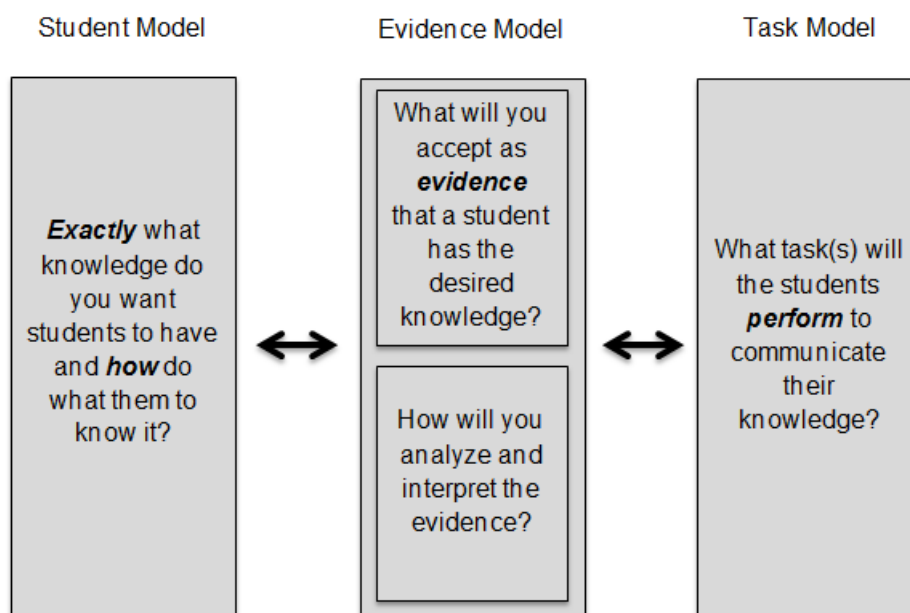


Figure 5.2. Simplified Representation of the Conceptual Assessment Framework of Evidence-Centered Design (Pellegrino et al., 2014)

This study provides examples of how students understand specific aspects of CATS items (tasks) and also provides description of the relationships between their performance and the forms of evidence (selection of distractors) that align with specific problematic reasoning. This interrelated nature of the ECD student model, evidence model and task model may also be viewed as projections of the assessment triangle in the same way that the findings from this study illustrate how these elements interrelate in the context of CATS. Continuing with ECD, this framework may be used to inform a design pattern for CATS which would allow for reproduction of CATS items for additional instructional purposes.

#### 5.4 Implications and Recommendations

Returning to the initial purpose of this dissertation study, it was expected that an exploratory approach to understanding CATS item difficulty would hold value for instructors using CATS by building upon the interpretation vertex of the assessment triangle. The following sections detail possible implications from the findings of this study, first by describing implications for instructors using CATS both as a formative and summative test. The second implications section focuses on what these findings may mean for other CI development and evaluation.

##### 5.4.1 Implications for CATS as an Instructional Tool

Even though CATS has performed well in previous structural evaluation such as factor analysis as a test that adheres well to a conceptual structure, instructors may consider additional factors beyond the targeted concept as potential influencers of student response patterns. Building off of the theory of difficulty for CATS: complexity of item content matters, as does complexity of item context. Models such as the conceptual

framework for statics may provide generalizable alignment of individual items with specific concepts, but instructors may consider the variety of problematic reasoning that can lead to incorrect responses. While previous studies have collected evidence that provide instructors using CATS valuable information regarding the targeted concepts for each item and trustworthiness that the test accurately measures mastery of specific skills, students' incorrect responses to individual items may not necessarily indicate a lack of mastery of the targeted concept. When using CATS as a formative test, instructors should consider whether additional instruction is needed to aid in students' performance with regard to the context of the items in addition to the content.

For example, students who incorrectly respond to CATS items grouped on Newton's 3rd Law (4, 5, 6) may not necessarily have difficulty with the concept of those items, but specific item features that contribute to the difficulty of the items. Explicit instruction that addresses force interactions of a pin joint that are separated for analysis both for an individual connection and for contexts involving frames with multiple connecting bodies and connections may aid in instruction that targets the contextual elements contributing to the difficulty of these items.

There is a need overall in instruction for awareness of inconsistencies in students' problematic reasoning and instructional practices that consider contextual elements of the domain in addition to principles. Examples of such approaches include a body-centric approach to statics problem solving (Steif, Lobue, Kara, & Fay, 2010) wherein improvements in solution accuracy are associated with induced talk about the conceptual relationships between bodies and the forces acting on them, as well as discussion-prompting conceptual item worksheets with feedback that has been found to improve test



scores on CATS (Steele, Brunhaver, & Sheppard, 2014). Instructors may also consider the use of individual CATS items or similar items as “clicker” questions to elicit discussion of conceptual and contextual aspects of statics principles. Item 27 may be a particularly useful item for discussion based on the variety of student reasoning observed in response to this item. For this same reason, item 27 may be less useful as a test item since it is unclear that the selection of specific distractors is indicative of any generalizable problematic reasoning.

#### 5.4.2 Implications for Concept Inventory Development

Following the general guidelines for concept inventory development, items that perform “well” are focused on specific concepts or misconceptions and can be solved without extensive mathematical problem solving. As described in chapter 2, concept inventory development traditionally follows an iterative process where items are developed through engagement with domain experts and pilot testing with students to ensure balance of specific concepts within a domain model. Findings from this study suggest that just as CIs are often developed with an interest in balancing specific concepts, test developers should also seek to balance item formats. Since the contextual aspects of CIs may contribute to item difficulty along with conceptual elements, a general recommendation is to use similar iterative methods to develop and evaluate item design formats and problem contexts as part of the CI item development. For example, if using student responses to open-ended items as a method of distractor generation, it may be useful to evaluate similar items with different item formats or different contexts to determine item task models that effectively target the desired conceptual reasoning. In essence this would be one way in which an ECD framework approach may be

implemented in CI development. Although further research is needed to develop a full design pattern for CATS, there is evidence that CIs can be reverse engineered with respect to an ECD design pattern as a means of revising an existing concept inventory or developing related items for instructional purposes (Denick et al., 2014).

This study may also serve as an example of the potential for qualitative research to provide context for psychometric measures, beyond confirmatory use, and to provide feedback into underlying conceptual framework. Assessment is always an imperfect measurement of students' actual knowledge in a domain; however through careful alignment of key assessment elements, errors in assessment may be mitigated. A guiding framework of this study stresses the importance of alignment among the assessment triangle elements. This may be achieved through iterative item and test development, and a comprehensive approach to validity that includes multiple forms of data collection and analysis. In addition to the use of qualitative studies as a means of generating distractor designs or confirming conceptual structure among test items, qualitative studies that explain psychometric measures such as item difficulty should be used for strengthening evidence of validity and informing the meaning of test scores. Findings from this and similar studies may not only provide context for why some items perform differently than others, but can also be leveraged to inform the conceptual basis of CIs and provide evidence as to how concept inventories can be improved.

#### 5.4.3 Future Research Opportunities

Throughout the preceding discussion of findings from this dissertation study, I have attempted to acknowledge the limitations of this study to inform upon all factors that may contribute to CATS item difficulty. A caveat of thematic analysis is the limited

perspective obtained through collected evidence; it is a methodological approach that does not intend to wholly explain the phenomenon of interest. Additionally, the thematic scheme I developed and applied to CATS items and students' verbal data may not be effective in teasing out all aspects of reasoning that explains students' test performance. An area in which there is obvious need for additional research is in students' problematic reasoning of couples and moments. Findings from this study suggest that students have difficulty understanding the contexts in which couples or moments may act and the independence of force and moment summations. In general, research on students' understanding of mechanics is more extensive on translational concepts than rotational.

Concept inventories are growing in use among higher education science and engineering; however it is still most common for these tests to serve as evaluations upon which instructional approaches are compared. CIs that have strong evidence of validity in both item mastery and errors can provide valuable information to instructors as diagnostic tools. There is a need for instructional practices that effectively address conceptual understanding of engineering principles and effective measures of conceptual understanding are critical to informing instruction. Evidence-Centered Design provides an opportunity to generate reproducible concept inventory items based on a detailed design pattern. Findings from this study may provide the start of a design pattern for CATS, but additional efforts are needed to build out a functional ECD framework for CATS. ECD is a potentially powerful method for CI development because it provides a structured design process and is a means of implementing the general rationale described by the assessment triangle. These efforts may also inform the development of concept

inventories in related domains such as dynamics and kinematics, which could then be used to measure students' conceptual progression along course sequences.

## REFERENCES

## REFERENCES

- Allen, K. (2006). *The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics*. (Doctor of Philosophy Dissertation), University of Oklahoma.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: SAGE Publications.
- Bransford, J. D., Brown, A. L., Cocking, R. R., Donovan, M. S., & Pellegrino, J. W. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, D.C.: National Academy Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.
- Brookes, D. T., & Etkina, E. (2009). "Force," ontology, and language. *Physical Review Special Topics - Physics Education Research, 5*(1), 010110.
- Case, S. M., & Swanson, D. B. (2001). *Constructing written test questions for the basic and clinical sciences*. Philadelphia, PA.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271-315.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

- Chi, M. T. H., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science*, 36(1), 1-61.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50(1), 66-71.
- Creswell, J. W. (2008). *Educational research : planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, N.J.: Pearson/Merrill Prentice Hall.
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38(2), 105-134.
- Denick, D. L., Miller, R. L., Streveler, R. A., Pellegrino, J. W., DiBello, L. V., Schroeder, L., . . . Jorion, N. (2014, Apr. 3-7). *The use of evidence centered design to reverse engineer an existing CI and develop new questions for diagnostic reporting*. Paper presented at the AERA, Philadelphia, PA.
- Denick, D. L., Santiago-Román, A. I., Pellegrino, J. W., Streveler, R. A., & DiBello, L. V. (2013). *Validating the diagnostic capability of the concept assessment tool for statics (CATS) with student think-alouds*. Paper presented at the ASEE Annual Conference & Exposition, Atlanta, GA.
- Denick, D. L., Santiago-Román, A. I., & Streveler, R. A. (2013, Apr. 27 - May 1). *Verbal analysis of students' conceptual reasoning*. Paper presented at the AERA, San Francisco, CA.

- Denick, D. L., Santiago-Román, A. I., Streveler, R. A., & Barrett, N. (2012). *Validating the diagnostic capabilities of concept inventories: Preliminary evidence from the concept assessment tool for statics (CATS)*. Paper presented at the 2012 ASEE Annual Conference & Exposition, San Antonio, TX.
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2/3), 105-225.
- diSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 35-60). New York: Routledge.
- Douglas, A. & Purzer, S. (2015). Validity: Meaning and relevancy in assessment for engineering education. *Journal of Engineering Education*. Guest Editorial Manuscript submitted for publication.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671-688.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, Mass.: MIT Press.
- Evans, D. L., Gray, G. L., Krause, S., Martin, J., Midkiff, C., Notaros, B. M., . . . Wage, K. (2003, Nov. 5-8). *Progress on concept inventory assessment tools*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Boulder, CO.



- Evans, D. L., Midkiff, C., Miller, R. L., Morgan, J., Krause, S., Martin, J., . . . Wage, K. (2002, Nov. 6-9). *Tools for assessing conceptual understanding in the engineering sciences*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Boston, MA.
- Gray, G. L., Costanzo, F., Evans, D., Cornwall, 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.
- Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Los Angeles: SAGE Publications.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.*, 66(1), 64.
- Haladyna, T. M. (2012). *Developing and validating multiple-choice test items*. New York, NY: Routledge.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, 64(10), 1316-1325.
- Hansen, M. A., & Steif, P. S. (2006, Apr. 7-11). *Using results from concept inventories to improve teaching: Providing feedback on concept scores*. Paper presented at the AERA Annual Meeting, San Francisco, CA.

- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *Phys. Teach.*, 30(3), 141-158.
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Jorion, N., James, K., Schroeder, L., & DiBello, L. V. (2013, Apr. 27 - May 1). *Statistical and diagnostic analyses of student performance on concept inventories*. Paper presented at the AERA, San Francisco, CA.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Krause, S., Decker, J. C., & Griffin, R. (2003, Nov. 5-8). *Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Boulder, CO.
- Krippendorff, K. (2013). *Content analysis: An introduction to its methodology*. Los Angeles; London: SAGE.
- Libarkin, J. C. (2008). *Concept inventories in higher education science: Manuscript prepared for the National Research Council*. Paper presented at the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, Washington, D.C.
- Lindell, R. S., Peak, E., & Foster, T. M. (2007, July 26-27). *Are they all created equal? A comparison of different concept inventory development methodologies*. Paper presented at the AIP Physics Education Research Conference, Syracuse, NY.

- Litzinger, T. A., Van Meter, P., Firetto, C. M., Passmore, L. J., Masters, C. B., Turns, S. R., . . . Zappe, S. E. (2010). A cognitive study of problem solving in statics. *Journal of Engineering Education*, 99(4), 337.
- Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.
- Mazur, E. (1997). *Peer instruction*. Upper Saddle River, N.J.: Prentice Hall.
- McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75-98). Hillsdale, N.J.: Erlbaum.
- McCracken, W. M., & Newstetter, W. C. (2001). *Text to diagram to symbol: Representational transformations in problem-solving*. Paper presented at the 31st Annual Frontiers in Education Conference, Reno, NV.
- Messick, S. (1990). *Validity of test interpretation and use*. Princeton, NJ: Educational Testing Service.
- Messick, S. (1994). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning*. Princeton, NJ: Educational Testing Service.
- Midkiff, K. C., Litzinger, T. A., & Evans, D. L. (2001). *Development of engineering thermodynamics concept inventory instruments*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Reno, NV.
- Minstrell, J. (2000). Student thinking and related assessment: Creating a facet-based learning environment *Grading the Nation's Report Card: Research from the Evaluation of NAEP* (pp. 44-73).

- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6-20.
- Montfort, D., Brown, S., & Findley, K. (2007). *Using interviews to identify student misconceptions in dynamics*. Paper presented at the 37th Annual Frontiers in Education Conference, Milwaukee, WI.
- Newcomer, J. L. (2010, October 27-30). *Inconsistencies in students' approaches to solving problems in Engineering Statics*. Paper presented at the 40th ASEE/IEEE Frontiers in Education Conference, Washington, DC.
- Newcomer, J. L., & Steif, P. S. (2006a). *Student explanations of answers to concept questions as a window into prior misconceptions*. Paper presented at the 36th ASEE/IEEE Frontiers in Education Conference, San Diego, CA.
- Newcomer, J. L., & Steif, P. S. (2006b). *Student explanations of answers to concept questions as a window into prior misconceptions*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, San Diego, CA.
- Newcomer, J. L., & Steif, P. S. (2008a). Student thinking about static equilibrium: Insights from written explanations to a concept question. *Journal of Engineering Education*, 97(4), 481-490.
- Newcomer, J. L., & Steif, P. S. (2008b, Oct. 22-25). *What students 'know' about statics: Specific difficulties common among students entering statics*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY.
- Newcomer, J. L., & Steif, P. S. (2008c). *What students 'know' about statics: Specific difficulties common among students entering statics*. Paper presented at the 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY.

- Olds, B. M., Moskal, B. M., & Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. *Journal of Engineering Education*, 94(1), 13-24.
- Ortiz, L. G., Heron, P. R. L., & Shaffer, P. S. (2005). Student understanding of static equilibrium: Predicting and accounting for balancing. *American Journal of Physics*, 73(6), 545.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.
- Pellegrino, J. W., DiBello, L. V., & Brophy, S. (2014). The Science and Design of Assessment in Engineering Education. In A. Johri & B. M. Olds (Eds.), *Cambridge Handbook of Engineering Education Research*. Cambridge: Cambridge University Press.
- Pellegrino, J. W., DiBello, L. V., Jorion, N., James, K., & Schroeder, L. (2013, Apr. 27 - May 1). *Components of a comprehensive approach to validity*. Paper presented at the AERA, San Francisco, CA.
- Prince, M., Vigeant, M., & Nottis, K. (2012). Development of the heat and energy concept inventory: Preliminary results on the prevalence and persistence of engineering students' misconceptions. *Journal of Engineering Education*, 101(3), 412-438.
- Reed-Rhoads, T., & Imbrie, P. K. (2008). *Concept Inventories in Engineering Education*. Paper presented at the National Research Council's Workshop Linking Evidence to Promising Practices in STEM Undergraduate Education, Washington, DC.

- Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive physics reasoning: A commitment to substance-based conceptions. *Cognition & Instruction, 18*(1), 1-34.
- Richardson, J. (2004, April). *Concept inventories: Tools for uncovering STEM students' misconceptions*. Paper presented at the Assessment and Education Research in the Course, Curriculum and Laboratory Improvement (CCLI).
- Rimoldini, L. G., & Singh, C. (2005). Student understanding of rotational and rolling motion concepts. *Physical Review Special Topics - Physics Education Research, 1*(010102), 1-23.
- Santiago-Román, A. I. (2009). *Fitting cognitive diagnostic assessment to the Concept Assessment Tool for Statics (CATS)*. (Engineering Education Ph.D.), Purdue University, West Lafayette, IN.
- Santiago-Román, A. I., Streveler, R. A., & DiBello, L. V. (2010, October 27-30). *The development of estimated cognitive attribute profiles for the Concept Assessment Tool for Statics*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Washington, DC.
- Santiago-Román, A. I., Streveler, R. A., Steif, P. S., & DiBello, L. (2010, June 20-23). *The development of a Q-matrix for the concept assessment tool for statics*. Paper presented at the ASEE Annual Conference and Exposition, Louisville, KY.
- Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: The National Academies Press.

- Stathopoulou, C., & Vosniadou, S. (2007). Conceptual change in physics and physics-related epistemological beliefs: A relationship under scrutiny. In S. Vosniadou, A. Baltas, & X. Vamvakoussi (Eds.), *Reframing the conceptual change approach in learning and instruction*. (pp. 145-163). New York, NY: Elsevier Science.
- Steele, K. M., Brunhaver, S. R., & Sheppard, S. D. (2014). Feedback from in-class worksheets and discussion improves performance on the Statics Concept Inventory. *International Journal of Engineering Education*, 30(4), 992-999.
- Steif, P. S. (2003, Nov. 5-8). *Comparison between performance on a concept inventory and solving of multifaceted problems*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Boulder, CO.
- Steif, P. S. (2004a, Oct. 20-23). *An articulation of the concepts and skills which underlie engineering statics*. Paper presented at the ASEE/IEEE Frontiers in Education, Savannah, GA.
- Steif, P. S. (2004b, June 20-23). *Initial data from a statics concept inventory*. Paper presented at the ASEE/IEEE Annual Conference and Exposition, Salt Lake City, UT.
- Steif, P. S. (2005, June 12-15). *Psychometric analysis of a statics concept inventory and its use as a formative assessment tool*. Paper presented at the ASEE Annual Conference and Exposition, Portland, OR.
- Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *Journal of Engineering Education*, 94(4), 363-371.

- Steif, P. S., & Hansen, M. A. (2006). Comparisons between performances in a statics concept inventory and course examinations. *International Journal of Engineering Education*, 22(5), 1070-1076.
- Steif, P. S., & Hansen, M. A. (2007). New practices for administering and analyzing the results of concept inventories. *Journal of Engineering Education*, 96(3), 205-212.
- Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving problem solving performance by inducing talk about salient problem features. *Journal of Engineering Education*, 99(Compendex), 135-142.
- Streveler, R. A., Brown, S., Herman, G., & Montfort, D. (2014). Conceptual change and misconceptions in engineering education: Curricular-, measurement-, and theoretically focused approaches. In A. Johri & B. M. Olds (Eds.), *Cambridge handbook of engineering education research*. Cambridge: Cambridge University Press.
- Streveler, R. A., Litzinger, T. A., Miller, R. L., & Steif, P. S. (2008). Learning conceptual knowledge in the engineering sciences: Overview and future research directions. *Journal of Engineering Education*, 97, 279-294.
- Streveler, R. A., Miller, R. L., Santiago-Román, 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.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.



Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169.

Vosniadou, S. (2002). On the nature of naïve physics. *Reconsidering conceptual change: Issues in theory and practice*, 61-76.

Wage, K. E., Buck, J. R., Wright, C. H. G., & Welch, T. B. (2005). The Signals and Systems Concept Inventory. *IEEE Transactions on Education*, 48(3), 448-461.

## APPENDICES

## Appendix A IRB Exemption



HUMAN RESEARCH PROTECTION PROGRAM  
INSTITUTIONAL REVIEW BOARDS

---

**To:** RUTH STREVELER  
ARMS

**From:** JEANNIE DICLEMENTI, Chair  
Social Science IRB

**Date:** 03/01/2012

**Committee Action:** Exemption Granted

**IRB Action Date:** 03/01/2012

**IRB Protocol #:** 1202011944

**Study Title:** COLLABORATIVE RESEARCH: Integrating Models of Cognition and Measurement with Concept Establishing the Validity and Diagnostic Capacity of Concept Inventories

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(1) .

If you wish to revise or amend the protocol, please submit a new exemption request. Please refer to our guidance "Minor Changes Not Requiring Review" on our website. Please contact our office if you have any questions.

We wish you good luck with your work. Please retain copy of this letter for your records.

Below is a list of best practices that you should be aware of and keep in mind when conducting your research.

**Category 1**

- Written permission from preschools, primary and/or secondary schools should be obtained prior to the Investigator engaging in research, such as recruitment and conducting research procedures. If the written permission was not submitted with the protocol at the time of IRB review (e.g., the school would not issue the letter without proof of IRB approval), the investigator must submit the written permission to the IRB office immediately upon receipt from the school. This is an institutional requirement.

**Categories 2 and 3**

- Surveys and data collection instruments should note that only participants 18 years of age and over are eligible to participate in the research, state that participation is voluntary and that any questions may be skipped, and include the investigator's name and contact information.
- Investigators should explain to participants the amount of time required to participate. Additionally, they should explain to participants how confidentiality will be maintained or if it will not be maintained.
- When conducting focus group research, investigators cannot guarantee that all participants in the focus group will maintain the confidentiality of other group participants. The investigator should make participants aware of this potential for breach of confidentiality.

- Written permission from businesses, preschools, primary and/or secondary schools should be obtained prior to the investigator engaging in research, such as recruitment and conducting research procedures. If the written permission was not submitted with the protocol at the time of IRB review (e.g., the school would not issue the letter without proof of IRB approval), the investigator must submit the written permission to the IRB office immediately upon receipt from the school. This is an institutional requirement.

#### Category 6

- Surveys and data collection instruments should note that participation is voluntary.
- Surveys and data collection instruments should note that participants may skip any questions.
- When taste testing foods which are highly allergenic (e.g., peanuts, milk, etc.) investigators should disclose the possibility of a reaction to potential subjects.

#### General

- To recruit from Purdue University classrooms, the instructor and all others associated with conduct of the course (e.g., teaching assistants) must not be present during announcement of the research opportunity or any recruitment activity. This may be accomplished by announcing, in advance, that class will either start later than usual or end earlier than usual so this activity may occur. It should be emphasized that attendance at the announcement and recruitment are voluntary and the student's attendance and enrollment decision will not be known by those administering the course.
- When conducting human subjects research at non-Purdue colleges and universities, investigators are urged to contact that institution's IRB to determine requirements for conducting research at that institution.
- When conducting human subjects research in places of business, investigators must obtain written permission from an appropriate authority from the business prior to engaging in research activities such as recruitment or conducting study procedures. This is an institutional requirement.

## Appendix B Interview Protocol

**Booklet 1 - Interview Protocol**

1. Begin the tape

2. State interview #, name of interviewee and interviewer(s), date

"During this session/interview/time, I will ask you to solve/answer some problems.

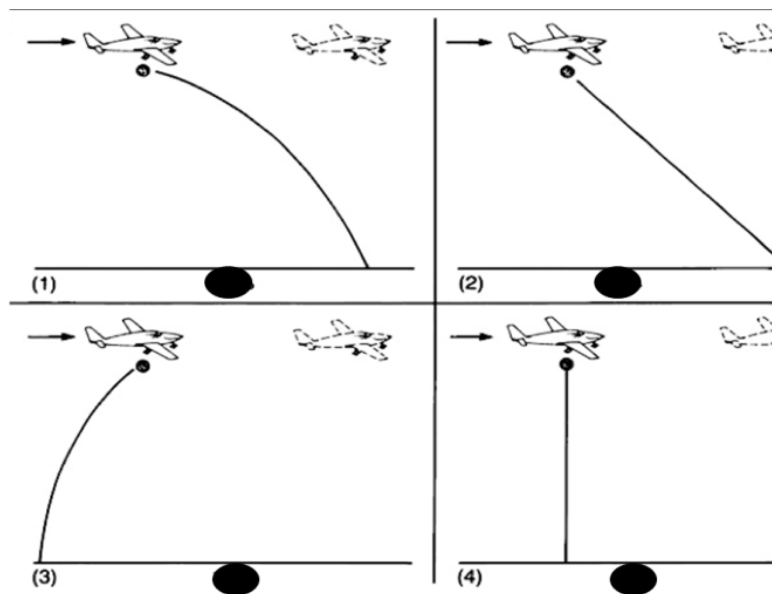
After you solve the problem, I will ask you to explain your answers. I may also ask you follow-up questions about your explanations. I am not looking for a particular answer and these problems are not meant to trip you up. I simply want to better understand your answers."

"The purpose of this interview is to evaluate the questions being asked, not your performance on the questions. Try to say as much as you can about what you're thinking as you work through the questions. Since we are recording the audio of this session, please read each question aloud and describe which answer you've chosen and why."

"Do you have any questions before we begin?"

"This first question that I'm going to give you is a warm up question, to get you used to explaining your answers. Please read the question aloud and then state your thoughts aloud as you solve the problem."

A plane is flying at a constant airspeed and drops an object. Which of the following is the correct representation of what happens to the dropped object?



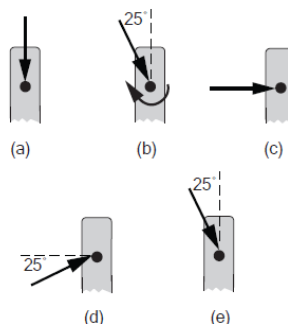
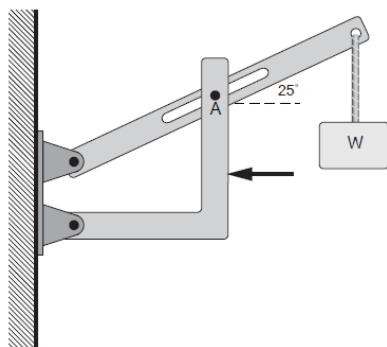
Please tell me why you selected response./Please explain your answer.

Please tell me why you did not select the other responses./Please explain why the other responses are incorrect.

3. Give the student the problems in the following order: 15, 14, 3, 1, 8, 7, 23, & 22.

15. The mechanism is acted upon by the horizontal force shown. The pin in the slot is frictionless.

What is the direction of the force exerted by the slot on the pin A?



### Question 15

Item #	SET	Concept	Skill01	Skill02	Slope	location
15	A	E	S05		0.859	-0.964

**Cognitive Attribute 10: Pin on slot** **Common Errors: 2, 3, 8, 12\*** (allows for moment at a point)

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what happens to the mechanism once the horizontal force acts. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Force is along member holding pin, (reactive force is perpendicular to applied force)

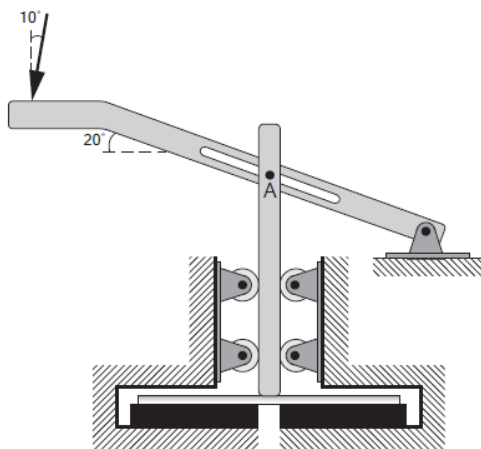
B: Reaction includes a couple (moment, torque, rotation)

C: Force acts to balance apparently an external force, (equal and opposite reaction)

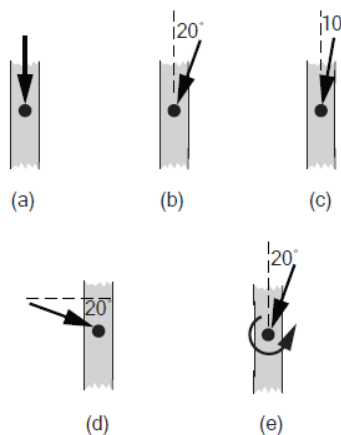
D: Force is parallel (in direction of ) to slot, (opposing reactive force)

**E: Correct: Force occurs at point of contact(at pin) and acts perpendicular(normal) to the slot**

14. The sheet metal punch is acted upon by the force shown acting at  $10^\circ$ . It drives the vertical ram which punches the sheet. The pin in the slot is frictionless.



What is the direction of the force exerted by the slot on the pin A?



### Question 14

Item #	SET	Concept	Skill01	Skill02	Slope	location
14	C	E	S05		0.958	-0.752

**Cognitive Attribute 10: Pin on slot** **Common Errors: 2, 3, 8, 12\*** (allows for moment at a point)

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what happens to the mechanism once the downward force acts. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Force is along member holding pin

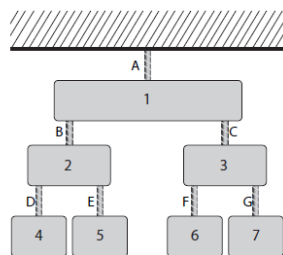
**B: Correct: Force occurs at point of contact(at pin) and acts perpendicular(normal) to the slot**

C: Force acts to balance apparently an external force, (force in direction of applied force)

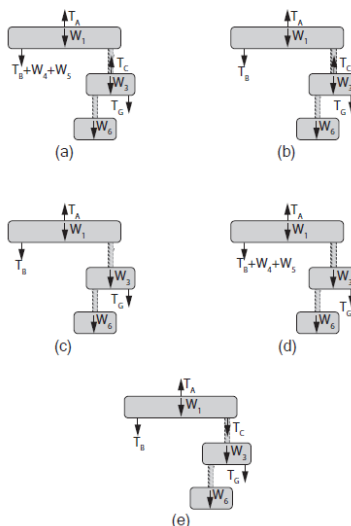
D: Force is parallel (in direction of ) to slot

E: Reaction includes a couple (moment, torque, rotation)

3. A free body diagram including blocks 1, 3, 6, and the cords connecting them, is to be drawn for this system.



Which is the correct free body diagram?



### Question 3

Item #	SET	Concept	Skill01	Skill02	Slope	location
3	C	A	S08		0.979	-0.327

**Cognitive Attribute 19 & 20: Representation**      **Common Errors: 1, 2, 3, 4, 5, 8**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you determined your answer choice.

#### Evidence of Steif's concepts:

A: Internal force; force plus weight (need to represent tension in included fbd cord upward, need to account for additional weight)

B: Internal force (need to represent tension in included fbd cord upward)

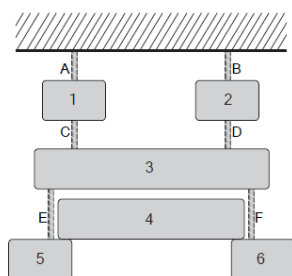
**C. Correct (fbd accounts for weights of included bodies, tensions indicated for bodies outside of system, tension not identified for bodies in system (internal forces))**

D: Force plus weight (need to account for additional weight)

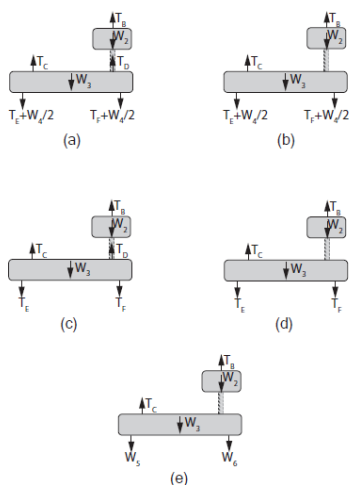
E: Internal force (need to represent tension in included fbd cord downward)



1. A free body diagram including blocks 2, 3 and the cord connecting them is to be drawn for this system.



Which is the correct free body diagram?



### Question 1

Item #	SET	Concept	Skill01	Skill02	Slope	location
1	A	A	S08		0.945	-0.071

**Cognitive Attribute 19 & 20: Representation** Common Errors: 1, 2, 3, 4, 5, 8

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you determined your answer choice.

#### Evidence of Steif's concepts:

A: Internal force; force plus weight (need to represent tension in included fbd cord upward, need to account for additional weight)

B: Force plus weight (need to account for additional weight)

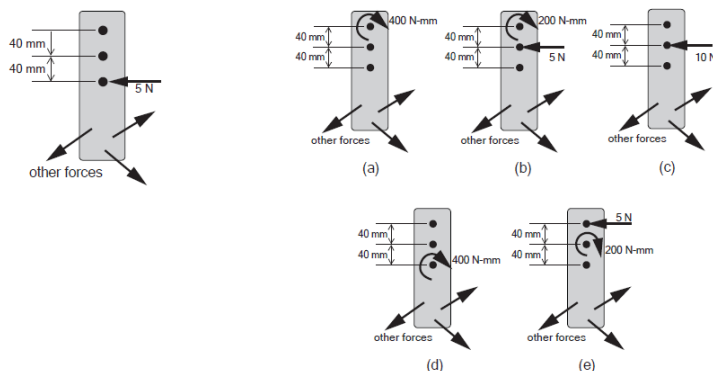
C: Internal force (need to represent tension in included fbd cord upward)

**D. Correct (fbd accounts for weights of included bodies, tensions indicated for bodies outside of system, tension not identified for bodies in system (internal forces))**

E: Weight instead of force (downward forces due to additional weight, not tension; tension only upward)

8. One force of magnitude 5 N keeps the member in equilibrium while it is subjected to other forces acting in the plane (shown schematically at the bottom). The three dots denote equally spaced points along the member.

Assuming the other forces stay the same, what load(s) could replace the 5 N force and maintain equilibrium?



### Question 8

Item #	SET	Concept	Skill01	Skill02	Slope	location
8	A	C	S01	S10	0.675	0.667

**Cognitive Attributes 1 & 8, Concept C: Static Equivalence (equal net force, equal net moment); Understanding of couples** Common Errors: 7, 10, 11

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you calculated/determined your answer choice.

#### Evidence of Steif's concepts:

A: Force is not equivalent, (400 N.mm divided by 80 mm provides equivalent 5 N force)

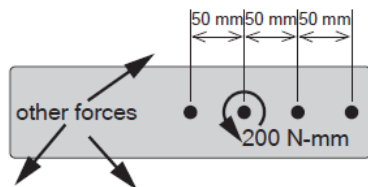
**B: Correct: Vector sum of forces equals zero; moment exerted is the same about any point; Force and moment are equivalent(balanced) - relocation of force causes a moment that must be balanced with an opposing moment; may be located at any point**

C: Force is not equivalent, (difference in force balanced by relocation of force, moments taken from top – 40 mm\*10 N = 80 mm\*5 N)

D: Force is not equivalent, (400 N.mm divided by 80 mm provides equivalent 5 N force)

E: Force is equivalent, but moment is not, (force is balanced, relocation of force is offset by location of moment)

7. A 200 N-mm couple acting counter-clockwise keeps the member in equilibrium while it is subjected to other forces acting in the plane (shown schematically at the left). The four dots denote equally spaced points along the member.



Item #	SET	Concept	Skill01	Skill02	Slope	location
7	A	C	S01	S10	0.734	0.793

**Cognitive Attributes 1 & 8, Concept C: Static Equivalence (equal net force, equal net moment); Understanding of couples** **Common Errors: 7, 10, 11**

Please tell me why you selected your response./Please explain your answer.

**Probes:**

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you calculated/determined your answer choice.

**Evidence of Steif's concepts:**

A: Force is equivalent(balanced), but moment is not - Evaluating from far right point:

$$100\text{mm} \cdot 200 \text{ Nmm} = 50\text{mm} \cdot 400\text{Nmm}$$

B: Force is not equivalent (balanced) - Evaluating from point with couple acting:

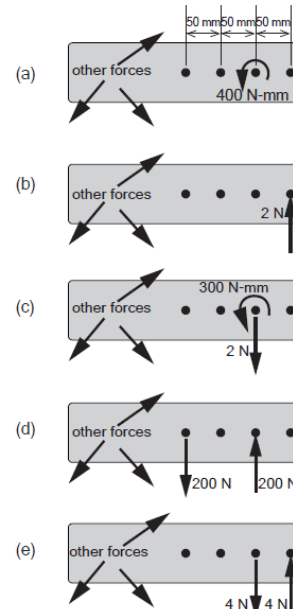
$$100\text{mm} \cdot 2\text{N} = 200\text{Nmm}$$

C: Force is not equivalent (balanced)

D: Force is equivalent(balanced), but moment is not

**E: Correct: Vector sum of forces equals zero; moment exerted is the same about any point; Force and moment are equivalent(balanced) - Students recognize that moment can be taken at any point - sum of moments at each point would equal 200Nmm**

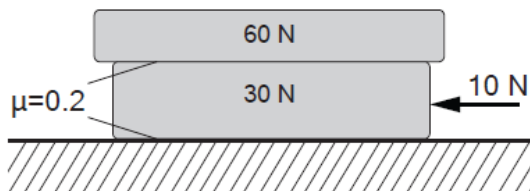
Assuming the other forces stay the same, what load(s) could replace the 200 N-mm couple and maintain equilibrium?



23. Two blocks are stacked on top of each other on the floor. The friction coefficient is 0.2 between all contacting surfaces (Take this to be both the static and kinetic coefficient of friction). Then, the horizontal 10 N force is applied to the lower block.

What is the horizontal component of the force exerted by the floor on the lower block?

- (a) 4 N      (b) 6 N      (c) 8 N      (d) 10 N      (e) 18 N



### Question 23

Item #	SET	Concept	Skill01	Skill02	Slope	location
23	A	H	S04		0.618	1.059

### Cognitive Attribute 7, Concept H: Friction Force Common Error: 9

Please tell me how you determined your answer/ Please tell me how you determined the number (depends on answer choice).

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Why is your calculation correct for this situation?

#### Evidence of Steif's concepts:

A: Normal force  $N$  is incorrectly computed ( $30\text{N} \times 0.2$ ); Friction force is difference between correct force (applied force) and  $\mu N$  ( $10\text{N} - 6\text{N}$ )

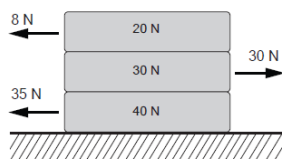
B: Normal force  $N$  is incorrectly computed ( $30\text{N} \times 0.2$ ); Friction force must equal  $\mu N$

C: Friction force is difference between correct force (applied force) and  $\mu N$

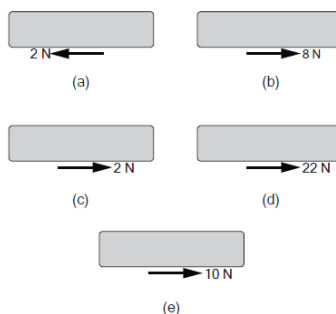
**D: Correct: Friction force is less than or equal to  $\mu N$  (static; applied force does not overcome frictional force, so Newton's 3rd law applies and resultant horizontal force is equal and opposite to applied force)**

E: Friction force must equal  $\mu N$  ( $90\text{N} \times 0.2$ )

22. Three blocks are stacked on top of one another on a table. Then, the horizontal forces shown are applied. The friction coefficient is 0.5 between all contacting surfaces. (This is both the static and kinetic coefficient of friction.)



Which of the following represents the horizontal component of the force acting on the lower face of the top (20 N) block?



### Question 22

Item #	SET	Concept	Skill01	Skill02	Slope	location
22	A	H	S04		0.473	1.249

### Cognitive Attribute 7, Concept H: Friction Force Common Error: 9

Please tell me how you determined your answer/ Please tell me how you determined the number (depends on answer choice).

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Why is your calculation correct for this situation?

#### Evidence of Steif's concepts: Misconception (Common Error)

A: Friction force is difference between correct force and  $\mu N$  ( $20 \text{ N} * 0.5 - 8 \text{ N} = 2\text{N}$ , in same direction as applied force)

**B: Correct: friction force acts in equal magnitude and opposite direction of applied force on bounded body)**

C: Friction force is difference between correct force and  $\mu N$  ( $20 \text{ N} * 0.5 - 8 \text{ N} = 2\text{N}$ , in opposite direction)

D: Null (friction force difference between applied forces  $30 \text{ N} - 8 \text{ N}$ ; unable to treat parts of a system as separate appropriately)

E. Friction force must equal  $\mu N$  ( $20 \text{ N} * 0.5 = 10 \text{ N}$ , in direction opposing applied force

Part 2- These questions will only be asked after all five questions have been covered with the participant if time permits.

*Are any of these questions familiar to you from your statics class?*

*Can you explain the hand gesture you made when solving question #?*

Question 15

5. Why should the force act 25 degrees from the vertical and not 25 degrees from the horizontal? (May point to D and E)
6. Why does/does not pin A not also have a moment? (May point to B)

Question 14

5. Why should the force act 20 degrees from the vertical and not 20 degrees from the horizontal? (May point to B and D)
6. Why does/does not pin A not also have a moment? (May point to E)
7. Why should the force 20 degrees from the vertical and not 10 degrees from the vertical? (May point to B and C)

Question 3

5. Why should/should not the left-most downward force include tension and/or weight in the free body diagram?
6. Why should/should not the tension in cord C be included in the free body diagram?

Question 1

5. Why should/should not the downward forces include tension and /or weight in the free body diagram?
6. Why should/should not the tension in cord D be included in the free body diagram?

Question 8

5. Why does/does not a moment at the top/bottom point maintain equilibrium? (May point at A or D)
6. Why does/does not a moment at the top/middle point and a 5N force at the middle/top point maintain equilibrium? (May point at B or E)
7. Why does/does not a 10N force at the middle point maintain equilibrium? (May point at C)

Question 7

5. Why does an upward load of 2N at the last point maintain equilibrium? (May point at B)
6. Why do a downward load of 200N at the first point and an upward load of 200N at the third point maintain equilibrium? (May point at D)
7. Why do a downward load of 4N at the third point and an upward load of 4N at the last point maintain equilibrium? (May point at E)

**Follow-up Questions:**

Did you find any of the questions ambiguous?

What grade did you receive in Statics?

What is your GPA?

Do you have any suggestions for the interview process?

## Booklet 2 - Interview Protocol

1. Begin the tape

2. State interview #, name of interviewee and interviewer(s), date

"During this session/interview/time, I will ask you to solve/answer some problems.

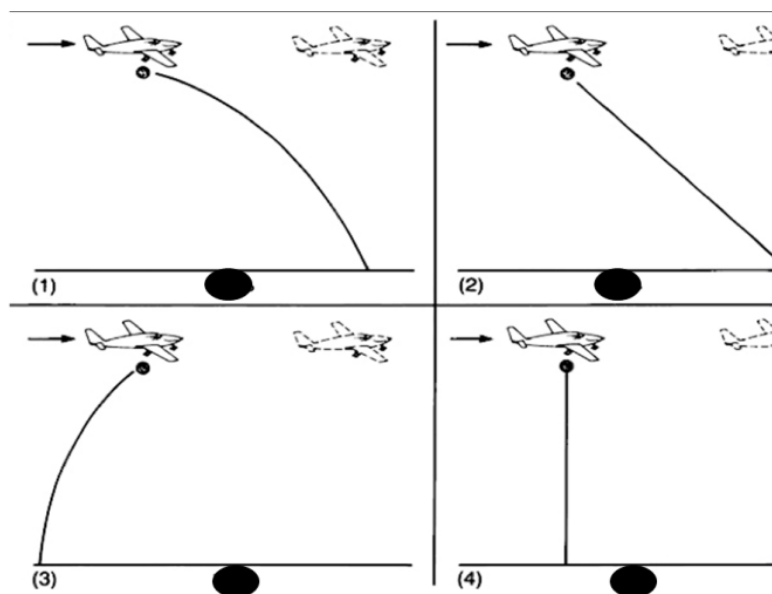
After you solve the problem, I will ask you to explain your answers. I may also ask you follow-up questions about your explanations. I am not looking for a particular answer and these problems are not meant to trip you up. I simply want to better understand your answers."

"The purpose of this interview is to evaluate the questions being asked, not your performance on the questions. Try to say as much as you can about what you're thinking as you work through the questions. Since we are recording the audio of this session, please read each question aloud and describe which answer you've chosen and why."

"Do you have any questions before we begin?"

"This first question that I'm going to give you is a warm up question, to get you used to explaining your answers. Please read the question aloud and then state your thoughts aloud as you solve the problem."

A plane is flying at a constant airspeed and drops an object. Which of the following is the correct representation of what happens to the dropped object?

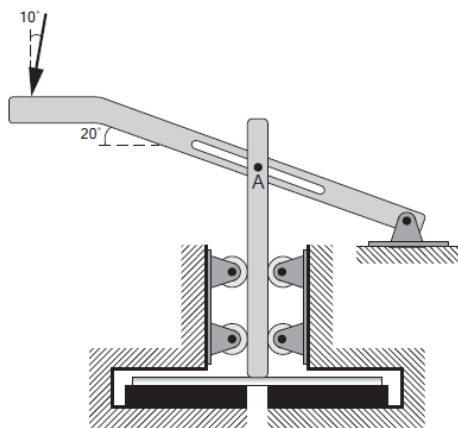


Please tell me why you selected response./Please explain your answer.

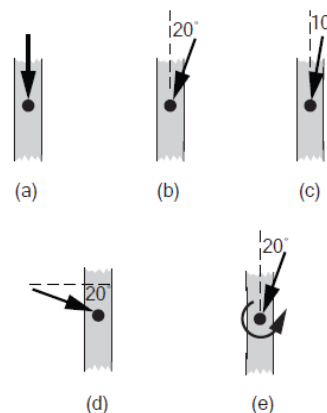
Please tell me why you did not select the other responses./Please explain why the other responses are incorrect.

3. Give the student the problems in the following order: 14, 19, 18, 3, 27, 5, 4, & 17.

14. The sheet metal punch is acted upon by the force shown acting at  $10^\circ$ . It drives the vertical ram which punches the sheet. The pin in the slot is frictionless.



What is the direction of the force exerted by the slot on the pin A?



### Question 14

Item #	SET	Concept	Skill01	Skill02	Slope	location
14	C	E	S05		0.958	-0.752

**Cognitive Attribute 10: Pin on slot** **Common Errors: 2, 3, 8, 12\*** (allows for moment at a point)

Please tell me why you selected your response./Please explain your answer.

#### Probes:

- I am not sure that I fully understand. Please give me some more detail/please tell me more.
- Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
- Were there any other responses that also appeared correct? Were there any other options that you were considering?
- Please walk me through what happens to the mechanism once the downward force acts. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Force along member holding pin

**B: Correct: Force occurs at point of contact(at pin) and acts perpendicular(normal) to the slot**

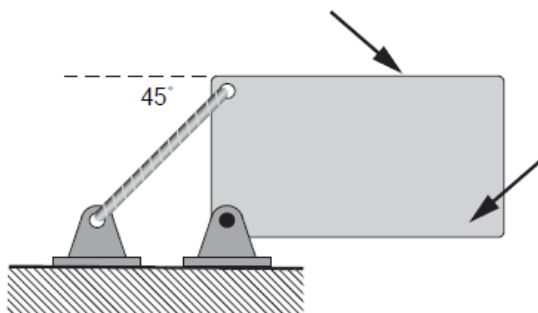
C: Force acts to balance apparently an external force, (force in direction of applied force)

D: Force parallel (in direction of ) to slot

E: Reaction includes a couple (moment, torque, rotation)

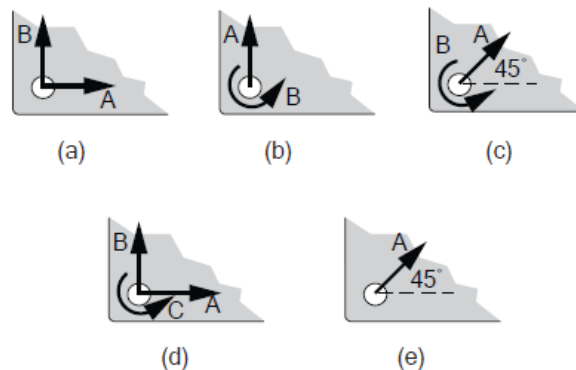


19. The plate could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin on the plate, as they would appear in a free body diagram of the plate. The pin is frictionless.



Which is the most appropriate partial free body diagram for the loads exerted by the pin on the plate in this situation?

The pin loads should be appropriate whatever the other forces on the plate.



### Question 19

Item #	SET	Concept	Skill01	Skill02	Slope	Location
19	B	G	S09		0.677	-0.574

**Cognitive Attribute 20: Representation of Forces** Common Errors: 1, 2, 3, 8

Please tell me why you selected your response./Please explain your answer.

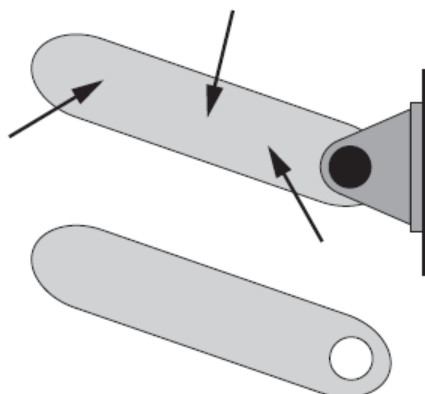
#### Probes:

8. I am not sure that I fully understand. Please give me some more detail/please tell me more.
9. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
10. Were there any other responses that also appeared correct? Were there any other options that you were considering?
11. Please explain how you determined your answer choice.

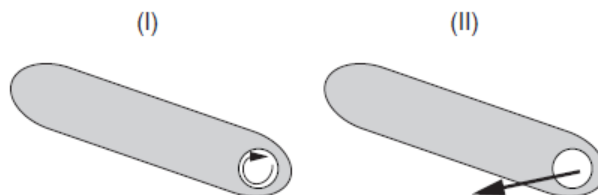
#### Evidence of Steif's concepts:

- A.
- B.
- C.
- D.
- E.

18. The arm can rotate freely about the pin. Additional loads act on the arm.



Could the reaction of the pin on the arm be as shown?



- (a) I is possible. II is possible  
 (b) I is possible. II is impossible  
 (c) I is impossible. II is possible  
 (d) I is impossible. II is impossible  
 (e) Cannot say without more information

### Question 18

Item #	SET	Concept	Skill01	Skill02	Slope	Location
18	B	F	S03	S09	0.581	-0.379

**Cognitive Attribute 5 & 20: Contact Forces & Representation of Forces** **Common Errors: 1, 2, 3, 8**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you determined your answer choice.

#### Evidence of Steif's concepts:

A: Couple may act (Reaction may include a moment)

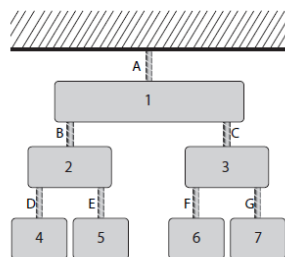
B: Couple may act (Reaction may include a moment)

**C: Correct - (Reaction may include a force outward from pin)**

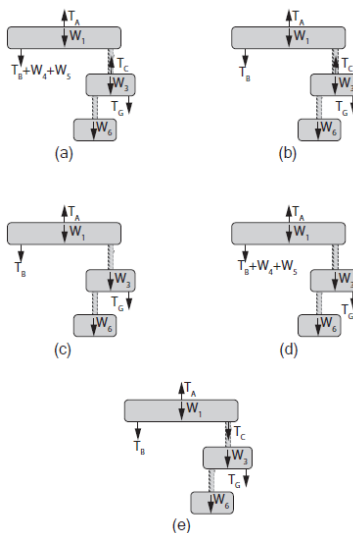
D: Null

E. Couple may act (Reaction may include a moment)

3. A free body diagram including blocks 1, 3, 6, and the cords connecting them, is to be drawn for this system.



Which is the correct free body diagram?



### Question 3

Item #	SET	Concept	Skill01	Skill02	Slope	location
3	C	A	S08		0.979	-0.327

**Cognitive Attribute 19 & 20: Representation**      **Common Errors: 1, 2, 3, 4, 5, 8**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please explain how you determined your answer choice.

#### Evidence of Steif's concepts:

A: Internal force; force plus weight (need to represent tension in included fbd cord upward, need to account for additional weight)

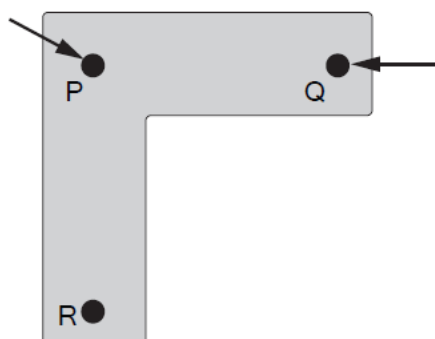
B: Internal force (need to represent tension in included fbd cord upward)

**C. Correct (fbd accounts for weights of included bodies, tensions indicated for bodies outside of system, tension not identified for bodies in system (internal forces))**

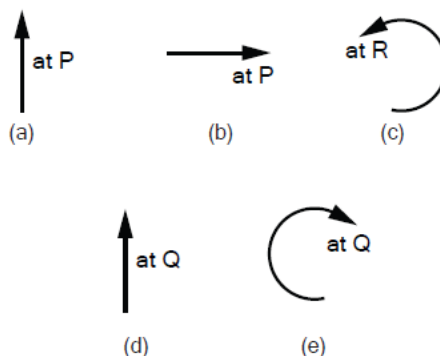
D: Force plus weight (need to account for additional weight)

E: Internal force (need to represent tension in included fbd cord downward)

27. The forces act at the points indicated. All magnitudes are greater than zero, and the forces act in the directions and senses shown.



Which one of the following additional loads, if given the right magnitude, could lead to equilibrium?



### Question 27

Item #	SET	Concept	Skill01	Skill02	Slope	Location
27	B	I	S10		0.635	0.048

**Cognitive Attributes 21: Couples & Equilibrium**    **Common Errors: 7, 10, 11**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what is happening to the object as the forces act at the points indicated. (May point to question diagram)

#### Evidence of Steif's concepts:

**A: Correct: Vector sum of forces equals zero (X and Y components of the forces, balanced forces; need for balancing vertical force), moment exerted is the same about any point**

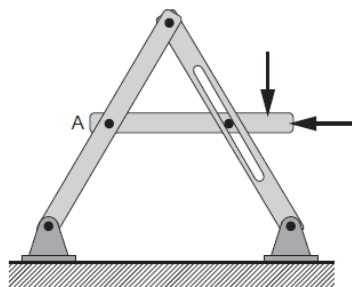
B: Forces need not be balanced (answer would result in downward vertical motion)

C: Forces need not be balanced (answer would result in downward vertical motion and moment, rotation)

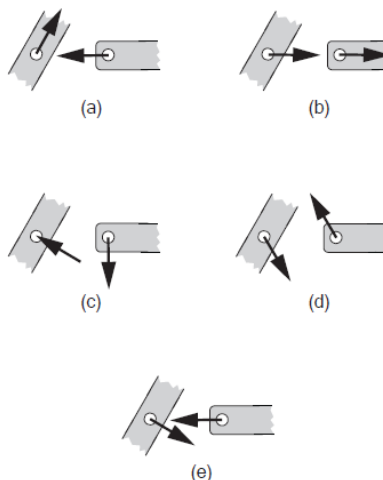
D: Moments need not be balanced (answer would result in moment, rotation)

E: Forces need not be balanced (answer would result in downward vertical motion and moment, rotation)

5. The frame shown could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin at A.



Which of the following could represent the forces exerted by the pin on the two bars? Forces act in the senses shown.



### Question 5

Item #	SET	Concept	Skill01	Skill02	Slope	Location
5	B	B	S02		0.469	0.354

**Cognitive Attribute 2: Newton's 3rd Law**

**Common Errors: 2, 6, 8**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what is happening to the object as the forces act at the points indicated. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Forces must be parallel to members

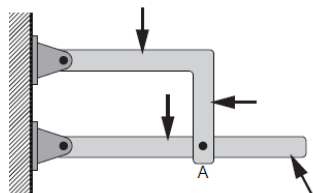
B: Null - (Forces must be reactive to applied force)

C: Forces must be perpendicular to members

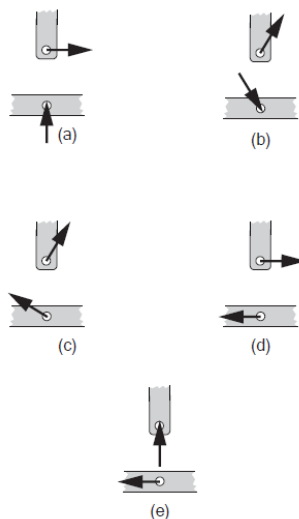
**D. Correct (Forces at pin must be equal and opposite to each other)**

E: Null

4. The frame shown could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin at A.



Which of the following could represent the forces exerted by the pin on the two bars? Forces act in the senses shown.



#### Question 4

Item #	SET	Concept	Skill01	Skill02	Slope	Location
4	B	B	S02		0.756	1.087

**Cognitive Attribute 2:** Newton's 3rd Law

**Common Errors:** 2, 6, 8

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what is happening to the object as the forces act at the points indicated. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Forces must be perpendicular to members

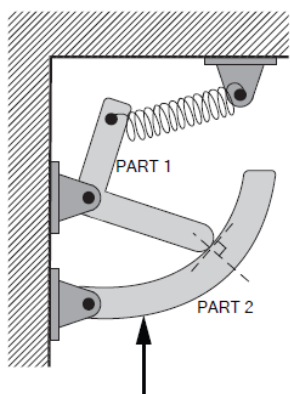
B: Null

C: Null

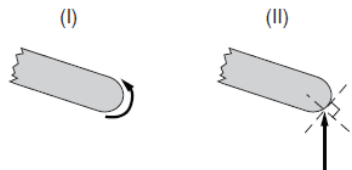
**D. Correct (Forces at pin must be equal and opposite to each other)**

E: Forces must be parallel to members

17. The circular portion of part 1 contacts the curved portion of part 2. Neglect friction between part 1 and part 2. Consider the bodies to be in the orientations shown.



Could the reaction of part 2 on part 1 be as shown?



- (a) I is possible, II is possible  
 (b) I is possible, II is impossible  
 (c) I is impossible, II is possible  
 (d) I is impossible, II is impossible  
 (e) Cannot say without more information

### Question 17

Item #	SET	Concept	Skill01	Skill02	Slope	Location
17	B	F	S03	S09	0.668	1.12

**Cognitive Attribute 5 & 20: Contact forces & Representation of Forces** **Common Error: 1,2,3,8,**

Please tell me why you selected your response./Please explain your answer.

#### Probes:

1. I am not sure that I fully understand. Please give me some more detail/please tell me more.
2. Please tell me why you did not select the other responses. (The student can go through and explain why the other responses were not chosen) Probe: Why do you think/believe that is not correct? What led you to believe/think that?
3. Were there any other responses that also appeared correct? Were there any other options that you were considering?
4. Please walk me through what happens to the mechanism once the upward force acts. (May point to question diagram)

#### Evidence of Steif's concepts:

A: Couple(moment, rotation, torque) may act; Force may be non-normal to surface

B: Couple(moment, rotation, torque) may act

C: Force may be non-normal to surface; Contact force parallel to applied force

**D: Correct: Force occurs at point of contact and perpendicular(normal) to the surface (resultant contact force)**

E: Couple(moment, rotation, torque) may act; Force may be non-normal to surface

Part 2- These questions will only be asked after all five questions have been covered with the participant if time permits.

*Are any of these questions familiar to you from your statics class?*

*Can you explain the hand gesture you made when solving question #?*

Question 14

5. Why should the force act 20 degrees from the vertical and not 20 degrees from the horizontal? (May point to B and D)
6. Why does/does not pin A not also have a moment? (May point to E)
7. Why should the force 20 degrees from the vertical and not 10 degrees from the vertical? (May point to B and C)

Question 18

5. Why is I possible/impossible? (May point to figure I)
6. Why is II possible/impossible? (May point to figure II)
7. What other information is needed to answer this question? (if E chosen)

Question 3

5. Why should/should not the left-most downward force include tension and/or weight in the free body diagram?
6. Why should/should not the tension in cord C be included in the free body diagram?

Question 27

5. Why would an upward directed load at point P lead to equilibrium? (May point at A)
6. Why would a moment at point R lead to equilibrium? (May point at C)
7. Why would an upward directed load at point Q lead to equilibrium? (May point at D)

Question 17

5. Why is I possible/impossible? (May point to figure I)
6. Why is II possible/impossible? (May point to figure II)
7. What other information is needed to answer this question? (if E chosen)

**Follow-up Questions:**

Did you find any of the questions ambiguous?

What grade did you receive in Statics?

What is your GPA?

Do you have any suggestions for the interview process?



## Appendix C Content Analysis

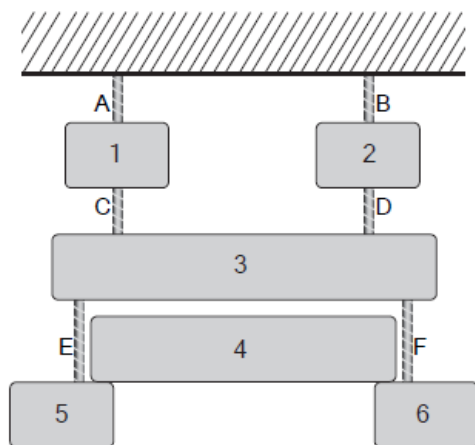
**Items 1 & 3**

**CMU Coding:** Concept A: Drawing forces on separate bodies; Concept Cluster 1, 3, 4

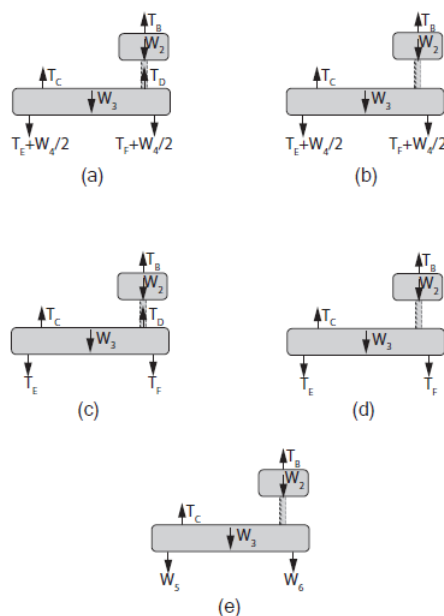
**Q-matrix Coding:** Cognitive Attribute 19 & 20: Representation & Tension in Ropes, Representation of Forces; Expected Common Errors: 1, 2, 3, 4, 5, 8

**Question 1 - Forces on collection of bodies**

1. A free body diagram including blocks 2, 3 and the cord connecting them is to be drawn for this system.



Which is the correct free body diagram?



**Direction of Problem Solving:** Forward

**Correct Response:** D

**Desired Reasoning:** Students are expected to identify the correct FBD by recognizing how force interactions are represented both inside and across system boundaries, recognizing that internal forces should not be included in an FBD (perhaps because both sides of force interaction are included within system as opposed to split across system boundaries), as well as recognizing when diagrams show redundant forces (Denick, Santiago-Román, Pellegrino, et al., 2013). Students need to recognize that tension acts where the cord attaches to block, not weight (Newcomer & Steif, 2006a).

**Expected Misconceptions:**

IntF: Internal force included improperly in FBD

W+F: force due to a connected cord improperly represented as the cord tension plus the weights of the non-attached bodies

WnotF: force due to a connected cord improperly represented as the weight of the proximal body only, rather than as the cord tension

(Steif & Hansen, 2007, p. 210)

Misconceptions:

- Internal Force (*E4*)
- Weight instead of force ( $\sim E5$ )
- Force plus weight ( $\sim E5$ )

1.

A: Internal force; force plus weight

B: force plus weight

C: Internal force

D. Correct

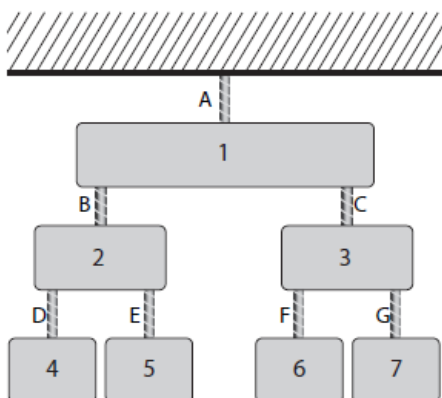
E: Weight instead of force

**Distractor Attractiveness:**

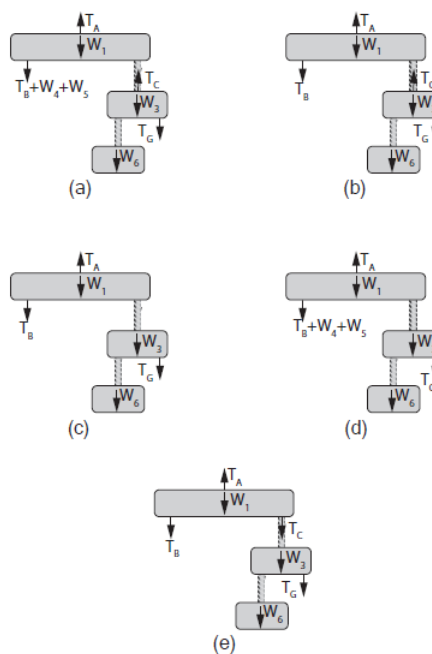
Not coded to thematic scheme. These items deal more with representation conventions for FBDs than physical reasoning.

**Question 3 - Forces on collection of bodies**

3. A free body diagram including blocks 1, 3, 6, and the cords connecting them, is to be drawn for this system.



Which is the correct free body diagram?



**Direction of Problem Solving:** Forward**Correct Response:** C

**Desired Reasoning:** Students are expected to identify the correct FBD by recognizing how force interactions are represented both inside and across system boundaries, recognizing that internal forces should not be included in an FBD (perhaps because both sides of force interaction are included within system as opposed to split across system boundaries), as well as recognizing when diagrams show redundant forces.

**Common Errors**

- 1 – Failure to be clear as to which body is being considered for equilibrium
- 2 – Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two
- 3 – Leaving a force off the FBD when it should be acting
- 4 – Drawing a force as acting on the body in the FBD, even though that force is exerted by a part which is also included in the FBD
- 5 – Drawing a force as acting on the body of the FBD, even though that force does not act directly on the body
- 8 – Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

Misconceptions:

- Internal Force (*E4*)
- Weight instead of force ( $\sim E5$ )
- Force plus weight ( $\sim E5$ )

3.

A: Internal force; force plus weight

B: Internal force

C: Correct

D: force plus weight

E: Internal force

**Distractor Attractiveness:**

Not coded to thematic scheme. These items deal more with representation conventions for FBDs than physical reasoning.

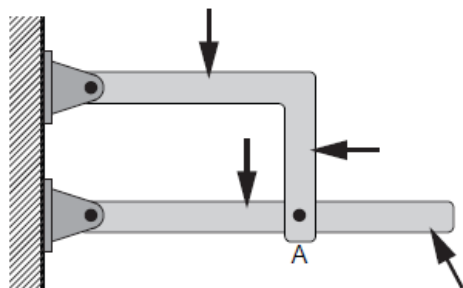
**Items 4 & 5**

**CMU Coding:** Concept B: Newton's 3<sup>rd</sup> Law; Concept Cluster: 1, 4

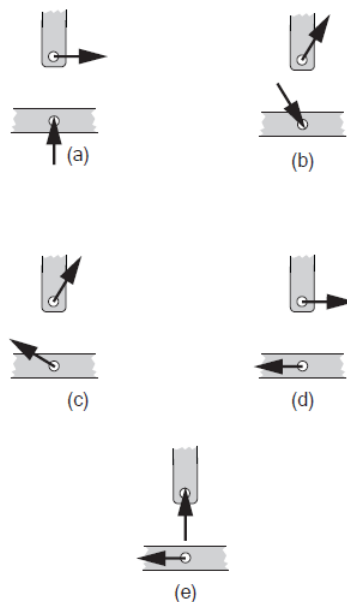
**Q-matrix Coding:** Cognitive Attribute 2: Newton's 3<sup>rd</sup> Law; Expected Common Errors: 2, 6, 8

**Question 4 - Newton's 3rd Law**

4. The frame shown could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin at A.



Which of the following could represent the forces exerted by the pin on the two bars? Forces act in the senses shown.



**Direction of Problem Solving:** Reverse

**Correct Response:** D

**Desired Reasoning:** Students are expected to recognize that when separating a body for analysis, any force pairs at a connection act in equal and opposite directions.

Students need to understand that the forces shown in the given diagram are representative of any forces that may be acting on a body and not the specific forces of interest for analysis.

**Expected Misconceptions:**

Misconceptions:

- Force must be parallel to member (*E6*)
- Force must be perpendicular to member (*E6*)

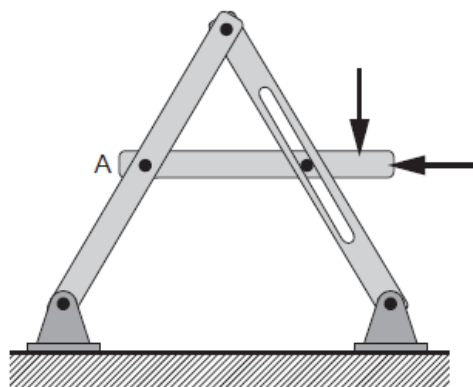
4.  
 A: Force must be perpendicular to member  
 B: Null  
 C: Null  
 D. Correct  
 E: Force must be parallel to member

#### Distractor Attractiveness

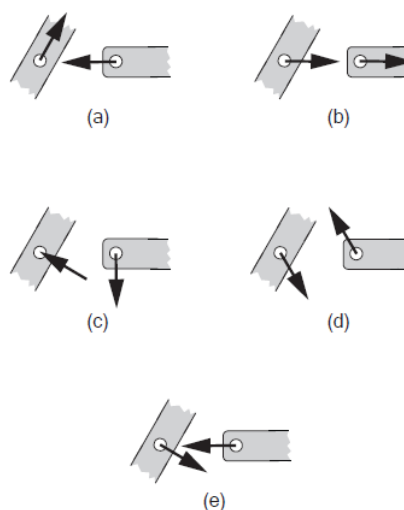
4.  
 A: Misapplication: Force perpendicular to member  
 B: Null  
 C: Null  
 D. Correct  
 E: Misapplication: Force along (parallel to) member

#### Question 5 - Newton's 3rd Law

5. The frame shown could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin at A.



Which of the following could represent the forces exerted by the pin on the two bars? Forces act in the senses shown.



**Direction of Problem Solving:** Reverse

**Correct Response:** D

**Desired Reasoning:** Students are expected to recognize that when separating a body for analysis, any force pairs at a connection act in equal and opposite directions.

Students need to understand that the forces shown in the given diagram are representative of any forces that may be acting on a body and not the specific forces of interest for analysis.

**Common Errors:**

2 – Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two

6 – Failing to account for the mutual (equal and opposite) nature of forces between connected bodies that are separated for analysis

8 – Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

Misconceptions:

- Force must be parallel to member (*E6*)
- Force must be perpendicular to member (*E6*)

5.

A: Force must be parallel to member

B: Null

C: Force must be perpendicular to member

D. Correct

E: Null

**Distractor Attractiveness**

5.

A: Misapplication: Force along (parallel to) member

B: Null

C: Misapplication: Force perpendicular to member

D. Correct

E: Null

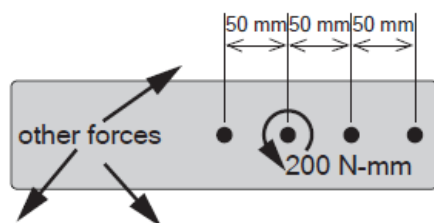
**Items 7 & 8**

**CMU Coding:** Concept C: Static Equivalence; Concept Cluster: 1, 2, 3, 4

**Q-matrix Coding:** Cognitive Attribute 1 & 21: Equivalence, Couples and Equilibrium;  
Expected Common Errors: 6, 7, 10, 11, 13\*

**Question 7 - Static Equivalence**

7. A 200 N-mm couple acting counter-clockwise keeps the member in equilibrium while it is subjected to other forces acting in the plane (shown schematically at the left). The four dots denote equally spaced points along the member.



**Direction of Problem Solving:** Reverse

**Correct Response:** E

**Desired Reasoning:** Students are expected to recognize that the given system is in equilibrium, (meaning that the sum of all forces and moments is zero), and recognize that the couple shown in the diagram could be replaced by a variety of forces that would produce the same moment and maintain the conditions of equilibrium.

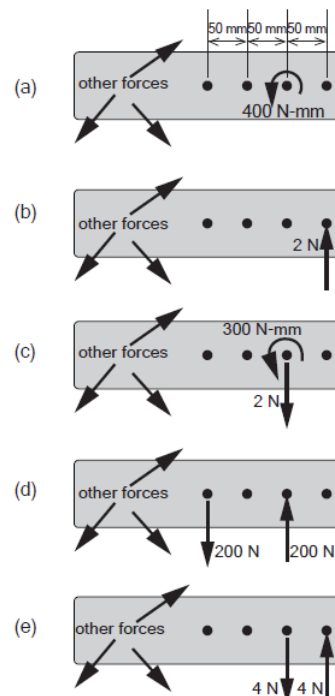
Students need to understand that the “other forces” in the diagram are intended to represent any variety of loads that in combination with the couple shown will result in equilibrium.

Students may need to understand how a couple results in a pure moment, and that while moments can be calculated about any point the location of a given couple is specific.

**Targeted Misconceptions:**

- Couples can only be broken into unbalanced forces
- Only couples can balance couples
- There are multiple axes of rotations

Assuming the other forces stay the same, what load(s) could replace the 200 N-mm couple and maintain equilibrium?



- Magnitudes of couples must change with location to have the same effect
- There is a specific point of rotation that must be maintained  
(Newcomer & Steif, 2008a)

**Expected Misconceptions:**

Misconceptions (*E10* includes multiple types of errors; and in some cases, *E11* is clearly not at issue, but in others, we cannot distinguish if *E11* or *E10* is source of student's difficulty with item – not satisfied with error classification of these items, or maybe items simply do not allow disentanglement of misconceptions

- Force is not equivalent (*E10* + *E11*)
- Force is equivalent, but moment is not (*E10* + *E11*)

7.

- A: Force is equivalent, but moment is not  
 B: Force is not equivalent  
 C: Force is not equivalent  
 D: Force is equivalent, but moment is not  
 E. Correct

**Distractor Attractiveness:**

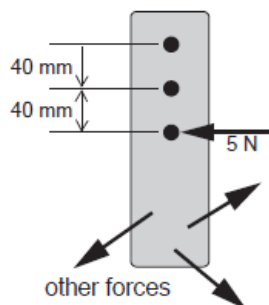
7.

- A: Partial: Force equivalent, not moment  
 B: Partial: Moment equivalent, not force  
 C: Partial: Moment equivalent, not force  
 D: Partial: Force equivalent, not moment  
 E. Correct

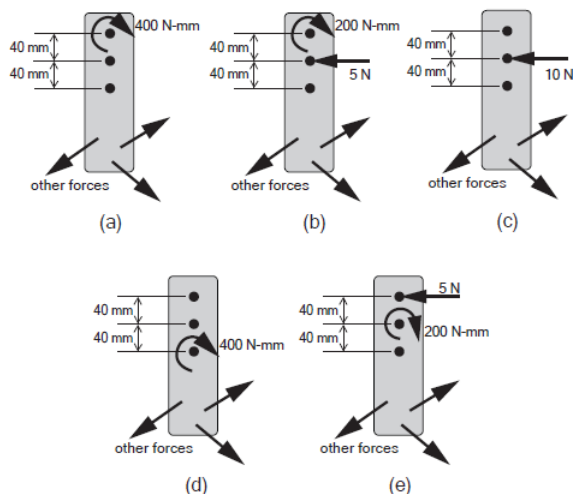
**Question 8 - Static Equivalence**



8. One force of magnitude 5 N keeps the member in equilibrium while it is subjected to other forces acting in the plane (shown schematically at the bottom). The three dots denote equally spaced points along the member.



Assuming the other forces stay the same, what load(s) could replace the 5 N force and maintain equilibrium?



**Direction of Problem Solving:** Reverse

**Correct Response:** B

**Desired Reasoning:** Students are expected to recognize that the given system is in equilibrium, (meaning that the sum of all forces and moments is zero), and recognize that the force shown in the diagram could be replaced by a variety of loads that would maintain the conditions of equilibrium.

Students need to understand that the “other forces” in the diagram are intended to represent any variety of loads that in combination with the 5 N force shown will result in equilibrium.

Students need to understand how the location of a force factors into a moment summation, and that while moments can be calculated about any point the location of a given couple is specific.

**Common Errors:**

6 – Failing to account for the mutual (equal and opposite) nature of forces between connected bodies that are separated for analysis

7 – Ignoring a couple that could act between two bodies or falsely presuming its presence.

10 – Failure to impose balance of forces in all directions and moments about all axes.

11 – Having a couple contribute to a force summation or improperly accounting for a couple in a moment summation

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

Misconceptions (*E10* includes multiple types of errors; and in some cases, *E11* is clearly not at issue, but in others, we cannot distinguish if *E11* or *E10* is source of student's difficulty with item – not satisfied with error classification of these items, or maybe items simply do not allow disentanglement of misconceptions

- Force is not equivalent (*E10* + *E11*)
- Force is equivalent, but moment is not (*E10* + *E11*)

8.

A: Force is not equivalent

B: Correct

C: Force is not equivalent

D: Force is not equivalent

E: Force is equivalent, but moment is not

**Distractor Attractiveness:**

8.

A: Partial: Moment equivalent, not force

B: Correct

C: Partial: Moment equivalent, not force

D: Partial: Moment equivalent, not force

E: Partial: Force equivalent, not moment

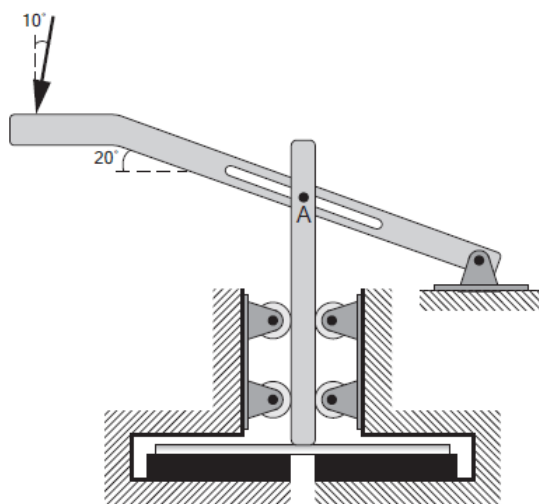
**Items 14 & 15**

**CMU Coding:** Concept E: Pin-in-slot Joint; Concept Cluster: 1, 3, 4

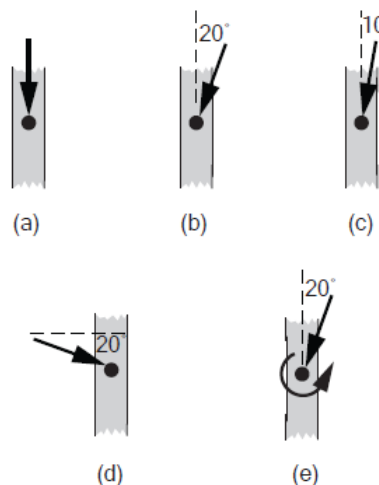
**Q-matrix Coding:** Cognitive Attribute 10: Pin in slot; Expected Common Errors: 2, 3, 8, 12\*

**Question 14 - Slot**

14. The sheet metal punch is acted upon by the force shown acting at  $10^\circ$ . It drives the vertical ram which punches the sheet. The pin in the slot is frictionless.



What is the direction of the force exerted by the slot on the pin A?



**Direction of Problem Solving:** Forward

**Correct Response:** B

**Desired Reasoning:** Students are expected to recognize that the connection shown at A is a pin-in-slot joint, and that the force pair between the two bodies within the system always acts in a direction perpendicular to the slot. Students may need to recognize that this is due to the degrees of freedom in other directions, namely in a translational direction along the slot and rotationally about the pin. Additionally, students may need to recognize that when the bodies are separated for analysis, conventional representation only shows one side of the acting force pair.

**Expected Misconceptions:**

Misconceptions:

- Force is parallel to slot (*E8*, similar to the errors in items 10-12)
- Force is along member holding pin (*E8*, similar to the errors in items 10-12)
- Reaction includes a couple (*E7*)

- Force acts to balance apparently an external force (E8, similar to the errors in items 10-12)

14.

A: Force along member holding pin

B: Correct

C: Force acts to balance apparently an external force

D: Force parallel to slot

E: Reaction includes a couple

**Distractor Attractiveness:**

14.

A: Misconception: Force along member

B: Correct

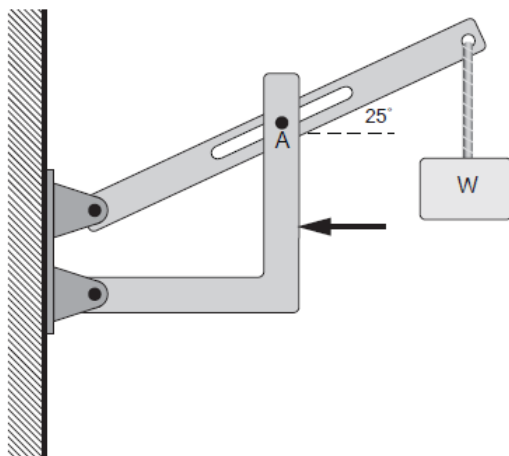
C: Misapplication: Equilibrium

D: Misconception: Force along member (slot)

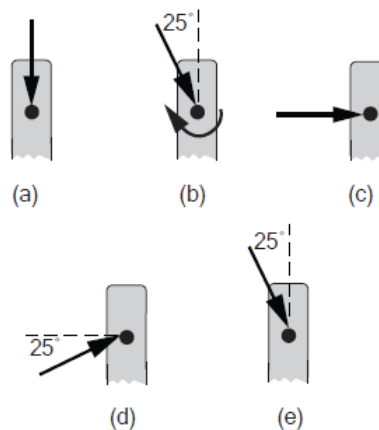
E: Misapplication: Moment

**Question 15 - Slot**

15. The mechanism is acted upon by the horizontal force shown. The pin in the slot is frictionless.



What is the direction of the force exerted by the slot on the pin A?

**Direction of Problem Solving:** Forward**Correct Response:** E

**Desired Reasoning:** Students are expected to recognize that the connection shown at A is a pin-in-slot joint, and that the force pair between the two bodies within the system always acts in a

direction perpendicular to the slot. Students may need to recognize that this is due to the degrees of freedom in other directions, namely in a translational direction along the slot and rotationally about the pin. Additionally, students may need to recognize that when the bodies are separated for analysis, conventional representation only shows one side of the acting force pair.

**Common Errors:**

2 – Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two

3 – Leaving a force off the FBD when it should be acting

8 – Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

Misconceptions:

- Force is parallel to slot (*E8*, similar to the errors in items 10-12)
- Force is along member holding pin (*E8*, similar to the errors in items 10-12)
- Reaction includes a couple (*E7*)
- Force acts to balance apparently an external force (*E8*, similar to the errors in items 10-12)

15.

A: Force along member holding pin

B: Reaction includes a couple

C: Force acts to balance apparently an external force

D: Force parallel to slot

E: Correct

**Distractor Attractiveness:**

15.

A: Misconception: Force along member

B: Misapplication: Moment

C: Misapplication: Equilibrium

D: Misconception: Force along member (slot)

E: Correct

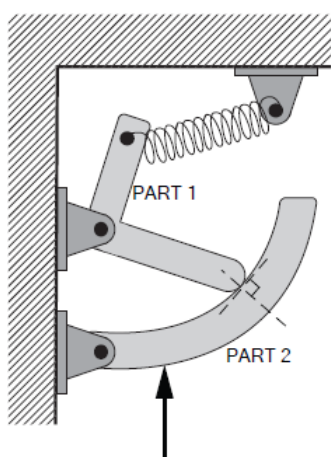
**Items 17 & 18**

**CMU Coding:** Concept F: Loads at surfaces with negligible friction; Concept Cluster: 1, 3, 4

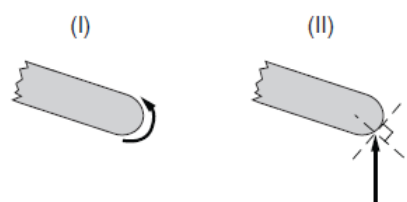
**Q-matrix Coding:** Cognitive Attribute 5 & 20: Contact Forces and Representation of Forces;  
Expected Common Errors: 1, 2, 3, 8, 12\*

**Question 17 - Negligible Friction**

17. The circular portion of part 1 contacts the curved portion of part 2. Neglect friction between part 1 and part 2. Consider the bodies to be in the orientations shown.



Could the reaction of part 2 on part 1 be as shown?



- (a) I is possible. II is possible  
 (b) I is possible. II is impossible  
 (c) I is impossible. II is possible  
 (d) I is impossible. II is impossible  
 (e) Cannot say without more information

**Direction of Problem Solving:** Complex

**Correct Response: D**

**Desired Reasoning:** Students are expected to recognize that the two curved bodies as shown in Part 1 and Part 2 of the diagram given are in contact at a point, and that a force interaction would act between the bodies. Students need to understand that the direction of the force interaction between the bodies would act in a direction perpendicular to the plane of contact. Students also need to recognize that each of the response options include at least a tangential force component, and are not possible for contact at a point.

**Expected Misconceptions:**

Misconceptions:

- Force may be non-normal to surface (*E8*)
- Couple may act (*E7*)

17.

A: Force may be non-normal to surface; Couple may act

B: Couple may act

C: Force may be non-normal to surface

D: Correct

E. Force may be non-normal to surface; Couple may act

**Distractor Attractiveness:**

17.

A: Misapplication: Moment; Misconception: Force in same direction as external force

B: Misapplication: Moment

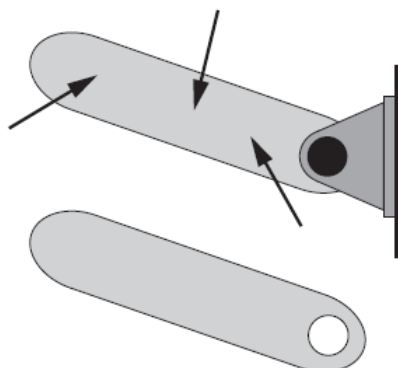
C: Misconception: Force in same direction as external force

D: Correct

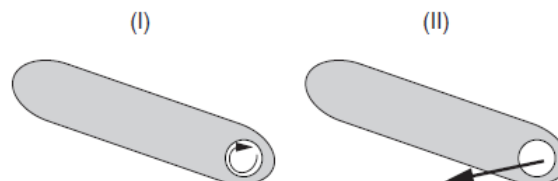
E. Null

**Question 18 - Negligible Friction**

18. The arm can rotate freely about the pin. Additional loads act on the arm.



Could the reaction of the pin on the arm be as shown?



(a) I is possible. II is possible

(b) I is possible. II is impossible

(c) I is impossible. II is possible

(d) I is impossible. II is impossible

(e) Cannot say without more information

**Direction of Problem Solving:** Complex**Correct Response:** C

**Desired Reasoning:** Students are expected to recognize that the arm is connected to a surface with a pin joint, and that there would be two perpendicular reaction forces. In order to understand that option II possible, students may need to understand that these reaction forces may be resolved as a single force in any direction. Students may need to recognize that this type of connection also includes a rotational degree of freedom about the pin.

Students need to understand that the forces shown in the given diagram are representative of any forces that may be acting on a body and not the specific forces of interest for analysis.

### Common Errors

- 1 – Failure to be clear as to which body is being considered for equilibrium
- 2 – Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two
- 3 – Leaving a force off the FBD when it should be acting
- 8 – Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces

(Santiago-Roman, 2009; Steif & Dantzer, 2005)

### Expected Misconceptions:

Misconceptions:

- Couple may act (*E7*)

18.

- A: Couple may act
- B: Couple may act
- C: Correct
- D: Null
- E. Couple may act

### Distractor Attractiveness:

18.

- A: Misapplication: Moment
- B: Misapplication: Moment
- C: Correct
- D: Null
- E. Null



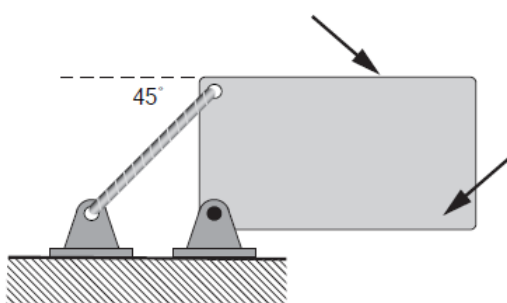
**Items 19**

**CMU Coding:** Concept G: Representing Loads at Connections; Concept Cluster: 1, 3, 4

**Q-matrix Coding:** Cognitive Attribute 20: Representation of Forces; Expected Common Errors: 1, 2, 3, 8, 12\*

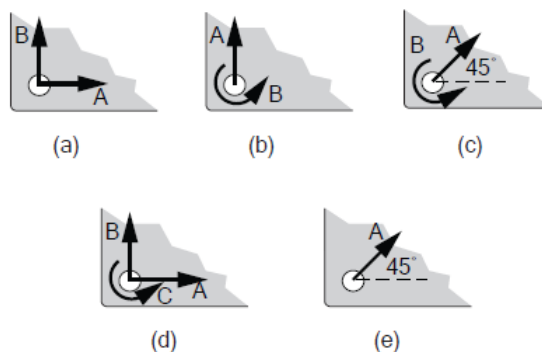
**Question 19 - Representation**

19. The plate could be subjected to a variety of forces applied at different points. Consider the loads exerted by the pin on the plate, as they would appear in a free body diagram of the plate. The pin is frictionless.



Which is the most appropriate partial free body diagram for the loads exerted by the pin on the plate in this situation?

The pin loads should be appropriate whatever the other forces on the plate.



**Direction of Problem Solving:** Forward

**Correct Response:** C

**Desired Reasoning:** Students are expected to recognize that the plate is connected to a surface with a pin joint, and that there would be two perpendicular reaction forces. Students may need to recognize that this type of connection also includes a rotational degree of freedom about the pin. Students may need to recognize that the angle of the cord connected to the plate adds to the variety of forces acting on the body, and does not affect the reaction forces at the pin (Denick, Santiago-Román, Pellegrino, et al., 2013).

Students need to understand that the forces shown in the given diagram are representative of any forces that may be acting on a body and not the specific forces of interest for analysis.

**Common Errors**

1 – Failure to be clear as to which body is being considered for equilibrium

2 – Failure to take advantage of the options of treating a collection of parts as a single body, dismembering a system into individual parts, or dividing a part into two

3 – Leaving a force off the FBD when it should be acting

8 – Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

Misconceptions:

- Couple may act (*E7*)
- Not allowing for the full range of possible forces between connected bodies, or not sufficiently restricting the possible forces (*E8*)

19.

A: Correct

B: *E7* and *E8*

C: *E7* and *E8*

D: *E7*

E. *E8*

**Expected Misconceptions:**

19.

A. Correct

B. Misapplication: Moment

C. Misapplication: Moment; Misapplication: Equilibrium

D. Misapplication: Moment

E. Misapplication: Equilibrium

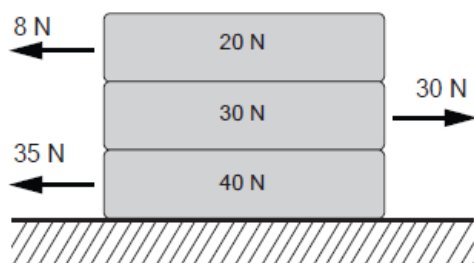
**Items 22 & 23**

**CMU Coding:** Concept H: Limits on Friction Force; Concept Cluster: 1, 3, 4

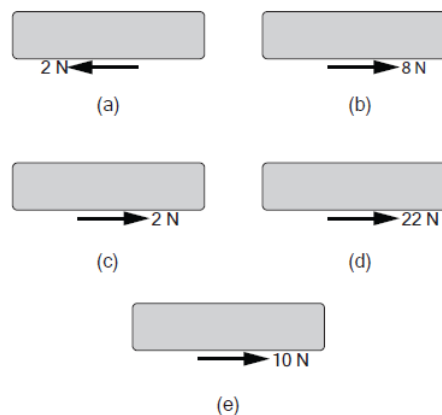
**Q-matrix Coding:** Cognitive Attribute 7: Friction Force; Expected Common Errors: 9

**Question 22 - Friction**

22. Three blocks are stacked on top of one another on a table. Then, the horizontal forces shown are applied. The friction coefficient is 0.5 between all contacting surfaces. (This is both the static and kinetic coefficient of friction.)



Which of the following represents the horizontal component of the force acting on the lower face of the top (20 N) block?



**Direction of Problem Solving:** Forward

**Correct Response:** B

**Desired Reasoning:** Students may need to recognize that the top block can be separated for analysis. By calculating the maximum static friction force for the top block ( $\mu N = (0.5)(20) = 10$ ), Students should recognize that the force applied to the top block is not large enough to set the top block into motion. Students need to understand that an equal and opposite force would describe the balancing friction force keeping the block in static equilibrium.

**Expected Misconceptions:**

- $\mu N$ : the tangential force is equal to the friction coefficient time the normal force even though a force of lesser magnitude maintains equilibrium
- $F - \mu N$ : the tangential force is equal to the applied force minus the friction coefficient times the normal force, even though a force of different magnitude maintains equilibrium (Steif & Hansen, 2007)

Misconceptions:

- Friction force must equal  $\mu N$  (E9)
- Normal force  $N$  is incorrectly computed (*probably implicit E3*)

- Friction force is difference between correct force and  $\mu N$  (*rare, unnamed variant of E9*)

22.

A: Friction force is difference between correct force and  $\mu N$ 

B: Correct

C: Friction force is difference between correct force and  $\mu N$ 

D: Null

E. Friction force must equal  $\mu N$ **Distractor Attractiveness:**

22.

A: Null

B: Correct

C: Null

D: Null

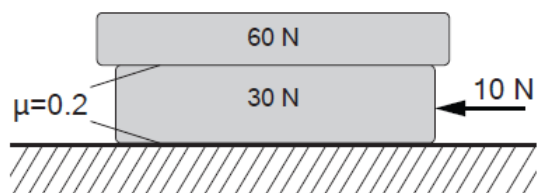
E. Misapplication: Friction

**Question 23 - Friction**

23. Two blocks are stacked on top of each other on the floor. The friction coefficient is 0.2 between all contacting surfaces (Take this to be both the static and kinetic coefficient of friction). Then, the horizontal 10 N force is applied to the lower block.

What is the horizontal component of the force exerted by the floor on the lower block?

- (a) 4 N      (b) 6 N      (c) 8 N      (d) 10 N      (e) 18 N

**Direction of Problem Solving:** Forward**Correct Response: D**

**Desired Reasoning:** Students may need to recognize that the blocks can be treated as a single body for analysis. By calculating the maximum static friction force for the blocks ( $\mu N = (0.2)(90) = 18$ ), Students should recognize that the force applied to the blocks is not large enough to set the blocks into motion. Students need to understand that an equal and opposite force would describe the balancing friction force keeping the blocks in static equilibrium.

### Common Errors

9 – Presuming a friction force is at the slipping limit ( $\mu N$ ), even though equilibrium is maintained with a friction force of lesser magnitude

#### Expected Misconceptions:

- $\mu N$ : the tangential force is equal to the friction coefficient times the normal force even though a force of lesser magnitude maintains equilibrium
- $F - \mu N$ : the tangential force is equal to the applied force minus the friction coefficient times the normal force, even though a force of different magnitude maintains equilibrium (Steif & Hansen, 2007)

Misconceptions:

- Friction force must equal  $\mu N$  (*E9*)
- Normal force  $N$  is incorrectly computed (*probably implicit E3*)
- Friction force is difference between correct force and  $\mu N$  (*rare, unnamed variant of E9*)

23.

A: Normal force  $N$  is incorrectly computed; Friction force is difference between correct force and  $\mu N$

B: Friction force must equal  $\mu N$ ; Normal force  $N$  is incorrectly computed

C: Friction force is difference between correct force and  $\mu N$

D: Correct

E. Friction force must equal  $\mu N$

#### Distractor Attractiveness:

22.

A: Null

B: Misapplication: Friction

C: Null

D: Correct

E. Misapplication: Friction

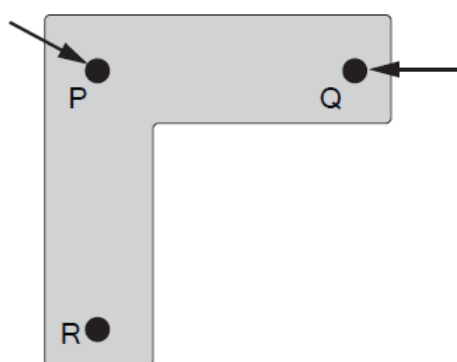
**Item 27**

**CMU Coding:** Concept I: Equilibrium; Concept Cluster: 1, 2, 3, 4

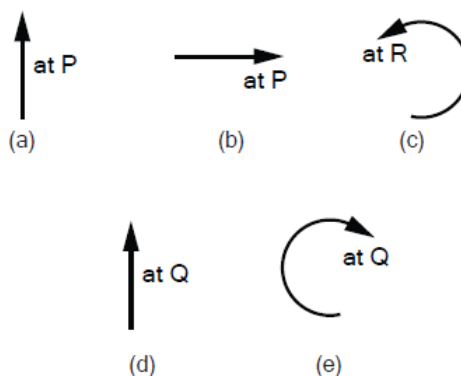
**Q-matrix Coding:** Cognitive Attribute 21: Couples & Equilibrium; Expected Common Errors: 7, 10, 11

**Question 27 - Equilibrium**

27. The forces act at the points indicated. All magnitudes are greater than zero, and the forces act in the directions and senses shown.



Which one of the following additional loads, if given the right magnitude, could lead to equilibrium?



**Direction of Problem Solving:** Reverse

**Correct Response:** A

**Desired Reasoning:** Students need to understand that static equilibrium requires that forces in the horizontal direction, forces in the vertical direction, and moments taken about a single point must all equal zero. Students need to understand that a vertical force (up) is needed to balance the vertical component (down) of the force acting on P. Students may need to understand that forces may cause rotation, but that moments or couples do not factor into force balances and that the moment summation should be zero for the diagram given.

Students need to recognize that the forces provided in the diagram should be analyzed in the directions provided, but that the magnitude of the forces is variable.

**Targeted Misconceptions:**

- Force equilibrium ignored selectively
- Force equilibrium always ignored
- Rotational equilibrium ignored selectively

- Rotational equilibrium always ignored
  - Couples include a force
  - Only couples balance couples
  - There are multiple axes of rotation
- (Newcomer & Steif, 2006a)

**Common Errors:**

7 – Ignoring a couple that could act between two bodies or falsely presuming its presence.

10 – Failure to impose balance of forces in all directions and moments about all axes.

11 – Having a couple contribute to a force summation or improperly accounting for a couple in a moment summation

(Santiago-Roman, 2009; Steif & Dantzler, 2005)

**Expected Misconceptions:**

- Forces need not be balanced ( $E_{I0} + E_{I1}$ )
- Moments need not be balanced ( $E_{I0} + E_{I1}$ )

27.

A: Correct

B: Forces need not be balanced ( $E_{I0}$ )

C: Forces need not be balanced ( $E_{I1}$ )

D: Moments need not be balanced ( $E_{I0}$ )

E: Forces need not be balanced ( $E_{I1}$ )

**Distractor Attractiveness:**

27.

A: Correct

B: Partial: Forces not balanced

C: Partial: Forces not balanced

D: Partial: Forces balanced, not moment

E: Partial: Forces not balanced

VITA



## VITA

Dana L. Denick received a Bachelor of Science degree in Mechanical Engineering and a Bachelor of Arts degree in Political Ecology from Bucknell University in 2002. She worked for six years as a Physics Teacher and Science Department Head at Esperanza Academy, a charter high school in Philadelphia, Pennsylvania. While teaching, Dana received a Master of Arts degree in Physics Education from the University of Virginia. Wanting to work with undergraduate engineering students, Dana then worked for two years as an Assistant Librarian for Engineering at Drexel University while completing a Master of Science degree in Library and Information Science. There, Dana began her engineering education research endeavors by developing a citation analysis assessment tool for undergraduate engineering design reports. Following this position, Dana worked as a Science Assistant at the National Science Foundation in the Division of Electrical, Communications, and Cyber Systems for one year.

Dana joined the School of Engineering Education at Purdue University in the Fall of 2011. Her research is focused on assessment development and validity, difficult concepts in engineering, and information literacy for engineering. While completing her dissertation, Dana accepted a position as Engineering/Science Analyst at the National Science Foundation in the Division of Civil, Mechanical, and Manufacturing Innovation. After graduating from the doctoral program, Dana will return to NSF and apply her

research experiences to supporting assessment and evaluation efforts of engineering research policy.