EXPLORING QUESTIONING AND COGNITION THROUGH THE MATHEMATICS MAJOR CURRICULUM

By

WILLIAM CHANCE BRADFORD, Bachelor of Science

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

Of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

December, 2015

Exploring Questioning and Cognition through the Mathematics Major Curriculum

By

WILLIAM CHANCE BRADFORD, Bachelor of Science

APPROVED:

Dr. Keith Hubbard, Thesis Director

Dr. Lesa Beverly, Committee Member

Dr. Crissy Cross, Committee Member

Dr. Kalanka Jayalath, Committee Member

Richard Berry D.M.A.

Dean of the Graduate School

ABSTRACT

Critical thinking has been identified by the Partnership for 21st Century Learning as one of the most necessary 21st century skills for students to succeed in the Information Age (Framework for 21st Century Learning, 2009). However, critical thinking is difficult for some to empirically define and measure. In order to investigate the development of critical thinking skills throughout the mathematics major, different classifications of cognition were researched and compiled into an instrument for investigating exam questions. Exams were collected from every course in the mathematics major at Stephen F. Austin State University (SFA), and the questions contained within them were classified according to eight different measurements of critical thinking. Trends in the required levels of critical thinking arose throughout the course sequence and correlations between the measurements were discovered.

ACKNOWLEDGEMENTS

I want to thank Dr. Hubbard for suggesting this topic for a thesis and for providing support every step of the way. I'd also like to thank the rest of my committee members, Dr. Beverly, Dr. Cross, and Dr. Jayalath for their time and hard work. Thank you to Dr. Harris and Dr. Stovall for providing valuable feedback. Thank you to all of the other professors I've had the pleasure of learning from throughout my education for helping me become the mathematician I'd like to be. I've had some great mentors help me along the way.

I'd like to thank all of the professors at SFA who let me analyze their exams. I appreciate your trust; this thesis wouldn't have been possible without it.

Thanks to my wife for tolerating the countless hours, late nights, and paperwork-filled weekends I've spent working towards my goals. Let's go on vacation for goodness' sake!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Critical thinking was defined by Michael Scriven and Richard Paul as "the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action" (Scriven & Paul, 1987). According to Richard Paul and Linda Elder (Paul & Elder, 2008), a well cultivated critical thinker

- raises vital questions and formulates them clearly and precisely;
- gathers and assesses relevant information, uses abstract ideas to interpret the information effectively;
- comes to well-reasoned conclusions and solutions, tests them against relevant criteria and standards;
- thinks open-mindedly within alternative systems of thought;
- assesses assumptions, implications, and practical consequences; and
- communicates effectively.

For many years there has been a call for teachers to cultivate their student's critical thinking ability, also known as their higher order cognitive skills (HOCS) (Paul, 1990; International Society for Technology in Education, 2002). HOCS can strengthen the brain, create more synapses between nerve cells, help learners make connections between past and new learning, and increase the likelihood that new learning will be consolidated and stored for future retrieval (Cardellichio & Field, 1997; Sousa, 2005). In 2015, a survey of 260 employers (Blue Cross and Blue Shield, Campbell Soup Company, ConAgra, Discover, Exxon, General Electric, Macy's, Proctor & Gamble, Progressive Insurance, Southwest Airlines, Toys "R" Us, Union Pacific Railroad Company, Verizon, etc.) rated Critical Thinking/Problem Solving as the most essential skill (National Association of Colleges and Employers, 2015).

However, studies have shown some college students are not well equipped with critical thinking skills. The researcher had difficulty in finding such studies specific to mathematics, but found studies conducted in other majors as well as across all majors. For example, in 1994 a statewide study was conducted in California to assess how well the state's teacher preparation programs were preparing elementary and secondary school teachers to teach critical thinking and problem-solving skills (Paul, Elder, & Bartell, 1997). Data were collected from 38 public colleges and universities and 28 private ones. The study found 89% of professors claimed critical thinking to be a primary objective of their instruction, but only 19% could give a clear explanation of what critical thinking is. In 2002, Bloom's Taxonomy co-author David Krathwohl noted "almost always these analyses (of curricular objectives and exam questions) have shown a heavy emphasis on objectives requiring only recognition or recall of information, objectives that fall in the Knowledge classification (of Bloom's Taxonomy)" (Krathwohl, 2002, p. 213). A national study in 2011 found 46% of college students (their majors were not specified) did not gain HOCS during their first two years of college, and 36% had not gained HOCS after four years (Arum & Roksa, 2011). In 2015, the Council for Aid to Education used the Collegiate Learning Assessment Plus (CLA+) to discover that out of 32,000 college students, 40% graduated without the complex reasoning skills (which included critical thinking) needed to manage white-collar work (Belkin, 2015).

In this study, final exams from every course in the mathematics major at Stephen F. Austin State University (SFA) were collected. The final exams were chosen for the analysis since they usually account for a large portion of the final grade, so professors most likely think the topics covered on the exams are comprehensive and representative of the entire course. The exams were analyzed in terms of higher-order critical thinking skills and questioning. Many different measurements of critical thinking were researched, modified, and combined to create a more comprehensive view of cognitive demands within the program. Detailed solutions of every exam were constructed, and measurements were applied to those solutions. The number of questions falling into each level of every classification were recorded. The purpose of this analysis is to examine the degree of critical thinking skills being tested throughout the undergraduate degree sequence.

CHAPTER 2

Literature Review

There is an abundance of information on how teachers can promote critical thinking skills during formative assessment (Black, 1998; Black & William, 2002; Connor-Greene, 2000; Fellenz, 2004; Hastings, 2003; Hannel, 2009). Studies show while teachers desire to promote HOCS, analyses of exams reveal mostly lower order cognitive skills (LOCS) are tested on exams (Black & Wiliam, 1998; Connor-Greene, 2000; Lemons & Lemons, 2013; Kokol-Voljc, 1999; Palmer & Devitt, 2007). While there are some measurements for analyzing HOCS (Bloom, 1956; Fisher-Hoch & Hughes, 1996; Kokol-Voljc, 1999; Lemons & Lemons, 2013; Stein & Smith, 2002), there is no single all-encompassing criteria, and no criteria describes all the steps needed to create a valid HOCS question in the first place (Lemons & Lemons, 2013). HOCS questions by definition must be novel to students and "typically include graphs, figures, case studies, or research designs" (Lemons, 2013). Of course, teachers must determine if students meet required course objectives. These requirements for novel questions meeting course objectives make it more difficult for instructors to create high quality HOCS questions. Designing questions that test HOCS can be especially difficult for some professors due to the concrete, factual, black-and-white nature of the subject (Karaali, 2011).

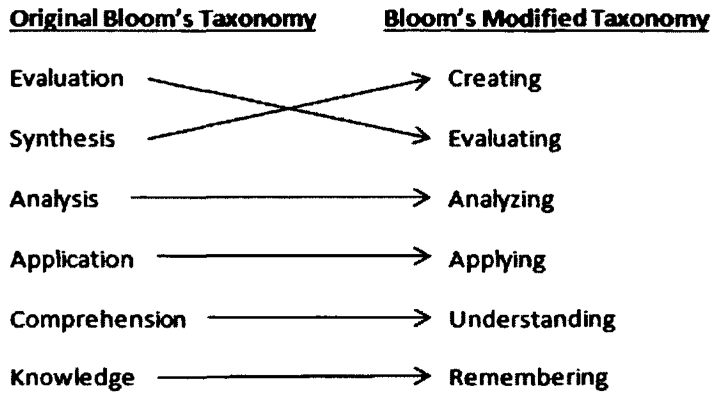
In order to measure the level of HOCS being tested on the exams, a list of eight classifications of cognitive demand was constructed by the researcher.

- Bloom's Modified Taxonomy
- Computer Algebra System (CAS)
- Concept/Visualization
- Language
- Question
- Stein & Smith (S&S)
- Steps Required
- Student Experience

Classifications

In 1956, Bloom's Taxonomy was created to "provide for classification the goals of our educational system", to help educators discuss questions/problems "with greater precision" (Bloom et al., p. 1). Bloom's Taxonomy was also designed to "promote the exchange of test materials and ideas about testing" and to stimulate "research on examining and on the relations between examining and education" (Bloom et al., p. 4). In 2000, Bloom's Taxonomy was modified slightly by Anderson and Krathwohl (Anderson et al., 2001). Figure 1 illustrates these changes.

Figure 1. Modifications to Bloom's Taxonomy



With so many categories in Bloom's Taxonomy, it can be difficult to classify a certain question. Therefore, the following compressed three level version of Bloom's taxonomy will be used, which is recommended by Palmer and Devitt (2007).

<u>Bloom's Modified Taxonomy Classification:</u>

1. Knowledge: Remember and Understand

   Recall data or information. Comprehend the meaning, state a problem in one's own words.

2. Comprehension and Application: Apply and Analyze

   Use a concept in a new situation. Separate concepts into component parts to understand organizational structure.

3. Problem-Solving: Evaluate and Create

   Make judgments about the values of ideas. Build a structure or pattern from diverse elements. Form parts into a whole, create new meaning.

Since its initial creation, Bloom's Taxonomy has been widely used in education to "classify curricular objectives and test items in order to show the breadth, or lack of breadth, of the objectives and items across the spectrum of (Bloom's Taxonomy) categories" (Krathwohl, 2002, p. 213). However, other measurements of gauging critical thinking are desired because Bloom's Taxonomy alone is inadequate (Lemons & Lemons, 2013).

Paula and Derrick Lemons found while Bloom's Taxonomy is indeed important in designing a challenging question, experienced professors who taught introductory biology courses (class sizes from 100-350 students per semester) at a large, public, Research I institution with more than 10 years experience each were unable to classify many questions using only Bloom's Taxonomy (Lemons, 2013). This led Lemons and Lemons to create three additional measurements: Difficulty, Time Required, and Student Experience.

The Difficulty ("whether a question was expected to be challenging for students", "whether the concepts referenced in a question are generally hard to apply", "how students typically perform on similar questions" (p. 51)) and Time Required ("whether the question could be answered quickly", "how many seconds or minutes a question could take" and "how many cognitive steps a student would go through to solve a question" (p. 53)) measurements are subjective. In order to solidify the levels of the

Difficulty classification, the researcher renamed this classification as Steps Required and accounted only for the number of significant cognitive steps needed to solve the problem. The author then removed the Time Required classification due to the similarity to the new classification Steps Required.

<u>Steps Required Classification:</u>

1. 1 to 2 significant cognitive steps
2. 3 to 4 significant cognitive steps
3. 5 or more significant cognitive steps

The student experience classification gauges the novelty of a question, which was selected as being a typical trait of a HOCS question by Lemons & Lemons (p. 48).

<u>Student Experience Classification:</u>

1. Routine –Student has surely seen a similar problem before, maybe many times.
2. Non-Routine – Student has seen similar questions, but not too similar or often.
3. Rare – Student has never or rarely worked with this or similar problems.

Fisher-Hoch and Hughes (1996) identified three categories of difficulty: Concept, Process, and Question. The Concept classification (renamed Concept/Visualization by the researcher to suit a wider range of applications) was based on "the intrinsic difficulty of the concept itself" (p. 2). This is a vague classification since difficulty is subjective. In an effort to establish a concrete measurement of difficulty, the researcher decided to classify a questions difficulty from the viewpoint of an average student with an overall A- to B+ average in the course.

Concept/Visualization Classification:

1. Problem can be easily visualized or conceptualized.
2. Problem contains some abstract ideas that may be difficult to conceptualize/visualize.
3. Problem contains some very abstract ideas that are very difficult to conceptualize/visualize.

The Process classification was defined as "the difficulty of cognitive operations and demands made on a candidate's cognitive resources" (p. 2). Cognitive resources were defined as "the student's re-construction of the meaning of the question", "the limitations of the student's working memory", and "irrelevant noise in the working memory" (p. 3). Since there was no possible way for the researcher to account for these differences among the individual students, this classification was omitted.

The Question classification measures the difficulty in understanding the phrasing of the question, subject matter aside. Formally, this classification gauges the level of valid/invalid (or intended/unintended) difficulty.

Question Classification:

1. A well-worded question with little possibility of multiple interpretations.
2. A question that is possibly confusing or open to multiple interpretations.
3. A question that is certainly confusing or open to multiple interpretations.

Finally, Fisher-Hoch and Hughes used previous work by Rothery (1980) to classify mathematical vocabulary words (symbols were not accounted for) by considering their meanings in "everyday" or "ordinary" English and mathematics.

Language Classification:

1. Words used in both English and mathematics that have roughly the same meaning (e.g. add, subtract, more, increase, intersect).
2. Words used in both English and mathematics but may have different meanings depending on context (e.g. area, volume, product, evaluate, differentiate, integrate, limit, group, ring).
3. Words used only in mathematics (e.g. hypotenuse, coefficient, matrix, quadratic, cubic, spline, subspace).

Another study categorized exam questions by making the assumption that the students have access to a Computer Algebra System (CAS) (Kokol-Voljc, 1999). This gave rise to the following categories that range from lowest order HOCS (1) to highest order HOCS (3). Students weren't actually allowed to use a computer on the exams. Instead, this classification system considers the difficulty of the question if the students were hypothetically allowed to utilize such a computer system. For this study, the level of CAS difficulty was determined by attempting to solve the question using the website Wolfram Alpha (WA). Since WA is very proficient in correctly interpreting input, questions classified as "Advanced" (requiring advanced knowledge of the CAS, or advanced syntax) were rare. There were many questions that required the student to do some preliminary work, but then could be solved by using WA simply as a calculator. In these cases, WA was not offering any more assistance than what could be found on a graphing calculator. For these reasons, this classification was changed from "Advanced" to "Advanced or Calculations" to both make up for the lack of "Advanced" questions and to give the calculation problems an appropriate place in the taxonomy.

Computer Algebra System (CAS) Classification:

1. Primary Routine – WA can do most of the work by simply typing in the question directly, or with only a few minor modifications.
2. Primary Advanced and Calculations – WA can do most of

the work, but advanced commands, clever wording, or preliminary work are required.

3. Secondary Routine – WA can do some of the work by simply typing in the question directly, or with only a few minor modifications.
4. Secondary Advanced and Calculations – WA can do some of the work, but advanced commands, clever wording, or preliminary work are required.
5. No CAS – WA offers absolutely no help.

Kokol-Voljc (1999) argued keeping the CAS system in mind when designing questions shifts the educational goals "from *performing* mathematical operations to *using* mathematical operations" (p. 1).

Stein and Smith created a four-level set of criteria for cognitive demand. This system was summarized well in an outline (Elliott & Stimpson, 2011), which has been further condensed for this analysis.

<u>Stein and Smith Classification:</u>

1. Memorization - Recall of facts; no algorithm is required; exact reproduction.
2. Processes Without Connections - Algorithmic, procedure is specifically stated or evident based on prior instruction or

experience. Focuses on correct answers rather than developing a mathematical understanding.

3. Processes With Connections - Thinking required. Procedure cannot be followed mindlessly. General procedures rather than algorithms must be followed. Questions may be accompanied by tables, graphs, illustrations, etc.

4. Doing Mathematics (Concepts and Processes) - Non-algorithmic activity. Requires access of relevant knowledge, self-reflection on actions, exploring concepts, analysis of constraints. Task is unpredictable due to nature of solution process required.

Before the results are analyzed, consider this quote from Kathleen Cotton:

> Quite a number of research studies have found higher cognitive questions superior to lower ones, many have found the opposite, and still others have found no difference...The conventional wisdom that says, "ask a higher level question, get a higher level answer," does not seem to hold (Cotton, 1988, p. 4).

The following analysis is not intended to label certain questions good and others not good, but to identify how the level of cognitive skills demanded by exam questions changes throughout the curriculum.

CHAPTER 3

Methodology for Analyzing the HOCS Demand of Final Exams

Exams were collected from every required credit-level course in the undergraduate mathematics degree sequence from the Fall 2013 and Spring 2014 semesters. Table 1 provides a list of these courses and their abbreviations (used in tables and figures).

Table 1

Abbreviations of Course Titles

Course Abbreviation	Course Title
PT	Plane Trigonometry
CA	College Algebra
PAG	Plane Analytic Geometry
PC	Pre-Calculus
C1	Calculus I
C2	Calculus II
NM	Introduction to Numerical Methods
MM	Introduction to Modern Math
LA	Linear Algebra
C3	Calculus III

DE	Differential Equations
CG	College Geometry
PM	Probability Modeling
SM	Statistical Modeling
AS	Algebraic Systems
A1	Introduction to Analysis I
A2	Introduction to Analysis II

All exams from the same course were grouped together to be analyzed as one. Any repeated questions were omitted. Then, detailed solutions were made for every question on all exams, and all significant cognitive steps were listed.

As an example of a significant cognitive step, the question "What is the slope of the line whose equation is $2y + 4 = 6x$?" was considered to have two cognitive steps. The first step is to realize the equation must be written in the form $y = mx + b$ to be able to identify the slope. The second step is to actually perform the algebra, subtracting 4 from both sides and then dividing the equation by 2. Note that both of these algebraic operations were counted as only one significant cognitive step (solving for y).

In dealing with questions with multiple parts, the whole question was classified as the highest rating among its parts. For example, if Part A was Bloom's level 1 and Part B was Bloom's level 2, the whole question was classified as a Bloom's level 2.

After every question was solved, it was assigned a rating from each HOCS measurement. Finally, average ratings for each classification were calculated for every exam. The previously mentioned eight classifications were then applied to every question in order to quantify the level of HOCS required. Note that questions used as examples have been modified to protect the confidentiality of the exams.

Organizing by Prerequisites

In order to examine the progression of HOCS measurements, it was necessary to create a consistent way to describe the difference between class levels. The choice was made to organize classes according to the number of prerequisite courses required. For example, Classification 0 courses required no prerequisites, whereas Classification 6 requires six prerequisites. Table 2 illustrates how the courses were divided into these six prerequisite categories.

Table 2

Prerequisite Categories

Prerequisite Classification	Courses in Classification
Classification 0	PT, CA
Classification 1	PAG, PC
Classification 2	C1
Classification 3	C2
Classification 4	NM, MM, C3, PM
Classification 5	LA, DE, CG, SM
Classification 6	AS, A1, A2

Other organizational schemas were considered, but ultimately rejected. In one of these schemas, courses were organized according to the SFA course codes. However, the scale of these ratings does not behave in a continuous way. For example, Calculus II (course MTH 234) is not somehow approximately 1.0043 times more difficult than Calculus I (course code MTH 233). In the other schema, courses were grouped according to the first number of the course code, which essentially distinguishes between Freshman (100 level), Sophomore (200 level), Junior (300 level), and Senior (400 level) courses.

Graphing the Data

A few additional changes were necessary in order to provide meaningful visual representations of the data. The data needed to be transformed to be on the same scale, and the weight of the questions needed to be considered.

First, the data were transformed to be on the same scale. For example, a question rated as the highest CAS rating of 5 is not 1.67 times more difficult than a question rated as the highest Bloom's level of 3. Instead, they should be considered as simply the highest level of cognitive demand. For this reason, all ratings were transformed to the scale of [0,1] where 0 is the lowest rating of the original classification system and 1 for the highest rating of the original classification system.

Initially, all questions were considered to have equal weight and all analyses were conducted under this assumption, but it was eventually discovered that the weight of the questions needed to be considered. For example, a professor most likely considers a question worth 20 points to be more important than a question worth only 2 points. Each question rating for every classification was changed using the following transformation (Equation 1) created by the researcher (these values will be referred to as the HOCS point transformations):

$$\frac{(Question\ Classification - 1)\ \times (Question\ Weight)}{(n-1)(Total\ Points\ on\ Exam)} \tag{1}$$

where

- *Question Classification* has possible values of 1, 2, 3, 4, or 5;
- *Question Weight* has a theoretical range of $(0, \infty)$,
- *n* is the highest level for the classification with possible values of 3, 4, or 5;
- *Total Points on Exam* has a theoretical range of $(0, \infty)$; and
- The range of the Point Transformation is [0,1].

CHAPTER 4

Results of Exam Analysis

Analysis of Percentages

First, the percentage of questions falling into each level of each classification (without considering question weight) was examined.

Table 3

Percentage of Questions in Each Level of HOCS Measurement

	Blooms			CAS					Steps		
	1	2	3	1	2	3	4	5	1	2	3
PT	0.19	0.69	0.13	0.25	0	0	0.38	0.38	0.06	0.44	0.50
CA	0.32	0.68	0	0.41	0.09	0.32	0	0.18	0.55	0.41	0.05
PAG	0.04	0.80	0.16	0.32	0.12	0.16	0.36	0.04	0	0.60	0.40
PC	0.09	0.68	0.24	0.32	0.09	0.15	0.44	0	0.06	0.26	0.68
C1	0.17	0.66	0.17	0.68	0.02	0.07	0.07	0.15	0.17	0.37	0.46
C2	0	0.78	0.22	0.22	0.11	0.22	0.33	0.11	0	0.33	0.67
NM	0	0.75	0.25	0	0	0	0.88	0.13	0.25	0.25	0.50
MM	0.24	0.33	0.42	0	0	0.03	0	0.97	0.33	0.36	0.30
LA	0.29	0.67	0.05	0	0	0.14	0.19	0.67	0.14	0.48	0.38
C3	0.20	0.60	0.20	0	0	0	0	[illegible]	0	0.40	0.60
DE	0.11	0.56	0.33	0	0	0	0.78	0.22	0	0	[illegible]
CG	0.08	0.54	0.38	0	0	0	0.31	0.69	0.08	0.08	0.85
PM	0.15	0.15	0.69	0	0	0	0.62	0.38	0	0.08	0.92
SM	0.20	0.10	0.70	0	0	0	0.40	0.60	0	0.20	0.80
AS	0.38	0.31	0.31	0	0	0	0.19	0.81	0.44	0.13	0.44
A1	0.50	0.30	0.20	0	0	0	0.30	0.70	0.20	0	0.80
A2	0.60	0.10	0.30	0.20	0	0	0.10	0.70	0.20	0.60	0.20

Table 3 Continued

	Concept 1	Concept 2	Concept 3	Stud Exp 1	Stud Exp 2	Stud Exp 3	Question 1	Question 2	Question 3
PT	.44	.44	0.13	69	0.31	0	[illegible]	0	0
CA	[illegible]	0	0	82	0.18	0	5	0.05	0
PAG	.52	0.32	0.16	84	0.32	0.04	8	0.04	0
PC	0.41	.47	0.12	68	0.32	0	4	0.06	0
C1	0.27	68	0.05	66	0.34	0	5	0.02	0.02
C2	0	78	0.22	0	[illegible]	0	89	0.11	0
NM	0	75	0.25	0	[illegible]	0	[illegible]	0	0
MM	0.36	61	0.03	.39	61	0	7	0.03	0
LA	0.24	71	0.05	.48	.52	0	[illegible]	0	0
C3	0	0.40	60	0.20	80	0	[illegible]	0	0
DE	0	0.44	56	0	[illegible]	0	[illegible]	0	0
CG	0	69	0.31	0.15	85	0	[illegible]	0	0
PM	0	69	0.31	.38	.46	0.15	[illegible]	0	0
SM	0	.50	.50	0.30	60	0.10	[illegible]	0	0
AS	0.06	75	0.19	56	0.31	0.13	[illegible]	0	0
A1	0.20	60	0.20	60	0.30	0.10	[illegible]	0	0
A2	0.20	80	0	80	0.20	0	[illegible]	0	0

Table 3 Continued

	Language 1	Language 2	Language 3	S&S 1	S&S 2	S&S 3	S&S 4
PT	0.19	63	0.19	0	.38	0.31	0.31
CA	.45	0.41	0.14	0.14	64	0.18	0.05
PAG	0.24	0.24	.52	0	0.20	68	0.12
PC	0.24	0.26	.50	0	0.35	.50	0.15
C1	0.20	0.15	66	0.05	0.15	68	0.12
C2	0.22	0.22	56	0	0.11	56	0.33
NM	0.13	0	88	0	.50	0.25	0.25
MM	0.18	0.09	73	0.24	0.09	0.12	55
LA	0.10	0.29	62	0.24	0.19	0.33	0.24
C3	0	0.40	60	0.20	0	60	0.20
DE	0	0	[illegible]	0	0	78	0.22
CG	0.15	0.08	77	0.08	0	.46	.46
PM	0.38	0	62	0.08	0	0.31	62
SM	0.30	0	70	0	0	.50	.50
AS	0	.50	.50	0.31	0.19	0.13	0.38
A1	0.10	0.10	80	.40	0.10	0.20	0.30
A2	0.30	0.20	50	.40	0.20	0.10	0.30

In the Bloom's Modified Taxonomy classification, Probability Modeling (PM) and Statistical Modeling (SM) stand out as having high percentages of level 3 questions. These exams contained many "real world" type problems that required the student to "build a structure or pattern from diverse elements" and set up the problem correctly before even attempting a solution. For example, consider the following question from Probability Modeling (the probabilities have been removed for confidentiality).

> A child playing hide and seek runs randomly in one of four directions when trying to find his hiding classmates: North, South, East, or West. If he runs North, he finds a classmate in the allotted time for the game N% of the time. Similar probabilities for running South, East, and West are S%, E%, and W%.
>
> A) What is the probability the seeking child will find a classmate in the time allotted for the game?
>
> B) If the seeking child finds a classmate, what is the probability the child ran South?

The question offers no hint as to what solution method to use. The Theorem of Total Probabilities and Bayes' Rule are essential in solving this problem ("build a structure or pattern from diverse elements" and "form parts into a whole"). To recognize this, not only must the student must have an understanding of these ideas, but recognize it's appropriate to apply them ("make judgments about the values of ideas").

Final exams in Algebraic Systems (AS), Introduction to Analysis I (A1), and Introduction to Analysis II (A2) had low average Bloom's scores due to a high number of definition questions.

In the CAS classification, Calculus I (C1) measured quite low. Those exams had many questions about calculating derivatives and limits. While these may be challenging for some students to solve by hand, WA solved them easily. There are a few different explanations for the classes with high CAS ratings. For proofs and "real-life scenario" type problems, WA is virtually useless. Without the student's ability to translate the relevant information from the problem into mathematics, there is very little WA can offer. As an example, consider the proof that the sum of two even numbers is even. While the idea behind the proof is relatively simple, WA offers no help in proving it.

In courses above Calculus II (C2), there was only one CAS I question asked. The question was an Introduction to Analysis II (A2) exam that asked students to evaluate an infinite series.

In the Steps classification (the number of significant cognitive steps required), College Algebra (CA) stood out with slightly over 50% of questions having only one to two steps. This is not surprising because many questions can be solved in one step by observation, such as those involving the degree of a polynomial, intercepts, asymptotes, domain and range.

In the Concept classification (which measures how difficult a question is to conceptualize/visualize), the Calculus III (C3) and Differential Equations (DE) final exams stand out with high percentages of level 3 questions because they deal with three dimensions. This is more challenging both conceptually and in terms of Calculus. To make matters more difficult for the students, they are often asked to set up integrals. Consider the following example:

> Set up an integral that would be used to find the surface area of $z = 2 - 3x^2 - 4y^2$ that lies above the xy-plane.

This is a challenging integral to set up because the student must have facility with three-dimensional visualization, integration in terms of a variable, and repeated integration.

Some of the senior level finals ranked low in the Concept classification due to many definition questions. These definitions may be quite challenging for some students to understand because every small detail is important. But the definition questions are not counted as conceptually difficult because the student is technically only recalling information. Although the Introduction to Analysis II (A2) final had no Concept 3 questions (containing very abstract ideas that are difficult to conceptualize/visualize), 80% were

Concept 2 (containing slightly abstract ideas that are somewhat difficult to conceptualize/visualize) , which was a higher percentage than all other exams. Perhaps one reason for there being no Concept 3 questions is that the students are never asked to create something "out of thin air" like on other exams. But as the high percentage of Concept 2 questions suggests, these questions are still challenging. Consider the definition of convergence of an infinite series, for which the students would need to write:

> Suppose that $\{a_n\}_{n=p}^{\infty}$ is a sequence of real numbers. We say that the series $\sum_{k=p}^{\infty} a_k$converges to some real value S if and only if the sequence of partial sums $\{S_n\}$ converges to S.

The idea behind this definition may be difficult for some students to grasp, but it is less conceptually difficult to write this definition than finding three dimensional surface areas.

Now consider the Student Experience classification (which measures how much experience the student has in solving a particular type of problem). The highest ranked final exams in this classification are Calculus II (C2), Introduction to Numerical Methods (NM), and Differential Equations (DE). Consider the following example of a Student Experience level 2 question from Calculus II (for confidentiality, the actual dimensions have been replaced by the variables m, x, y, and z):

> A trough has vertical ends that are trapezoids with parallel sides of length x ft. (top) and y ft. (bottom) and a height of z ft. The trough is filled with water to a depth of m ft. Set up a definite integral that represents the force exerted by the water on one end of the trough (density of water is $62.4 \text{ lbs}/\text{ft}^2$).

The students have surely seen a similarly structured problem before; otherwise it wouldn't be on the exam. However, it seems safe to assume this type of problem is slightly rarer than a simple integration problem with no connection to "real-life".

Now consider the Question classification (which measures valid vs. invalid difficulty, or unintended vs. intended difficulty). So few questions were ranked higher than level 1 that this classification was of little comparative value. In other words, there was very little invalid/unintended difficulty in the exam questions. The vast majority of the questions were clear and well-worded, leaving low possibility of confusion or multiple interpretations. For these reasons, the Question classification is largely omitted from the analyses. The one question rated as level 3 is discussed in the Correlation section of this research.

The upper level courses were also examined to see what percentage of the questions were proof-based (see Table 4). The final exam with the highest percentage of proofs was Introduction to

Modern Math. This isn't surprising because it's the course in which students first learn proof techniques.

Table 4

Percentage of Proof-Based Exam Questions

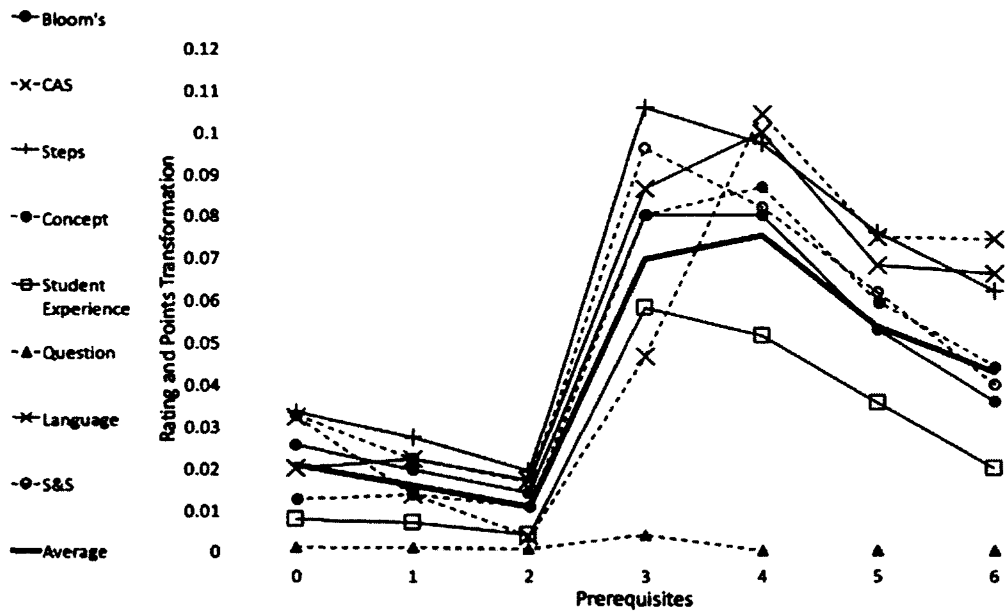
Class	Number of Proofs	Total Questions	Percentage of Proof-Based Questions
MM	13	30	43.33%
LA	3	24	12.50%
DE	1	9	11.11%
CG	4	11	36.36%
AS	4	15	26.67%
A1	1	10	10.00%

Note. Calculus III (C3), Probability Modeling (PM), Statistical Modeling (SM), and Introduction to Analysis II (A2) had no proofs.

Analysis of HOCS Transformations

The HOCS transformations were calculated using Equation (1). The average point values for each HOCS measurement were calculated for each exam, and then for each prerequisite level. Figure 3 gives a visual representation of these transformations.

Figure 2. Prerequisites vs. HOCS Transformations



Note that the baseline (Point Transformation = 0) represents the lowest rating possible for every HOCS classification, and Point Transformation = 1 (which is not shown on this scale) represents the highest rating possible.

The level of cognitive demand seemed to significantly increase from prerequisites levels 2 (Calculus I) and 3 (Calculus II). Perhaps this represents the first stages of students being asked questions requiring them to actually construct the problems they will need to solve. For example, consider this typical Calculus I problem:

The height of a ball (in meters) thrown into the air is given by $h(t) = -16t^2 + 60t + 6$, where *t* is time in seconds.

A) Find the vertical velocity at *t* = 1 second.

B) When will the ball reach its highest point?

C) How fast will the ball be falling when it hits the ground?

All of these questions involve finding the derivative (fairly algorithmic) with a few additional steps. The answer to part A is h'(1) = -32(1) + 60 = 28 m/s. Now consider a Calculus II problem:

Find the area enclosed by the graphs $y = x^2$ and $y = \sqrt{x}$.

The student must realize integration is necessary, recognize the upper function is $y = \sqrt{x}$, realize the bounds of integration are 0 and 1, set up the integral as

$$\int_0^1 \sqrt{x}\, dx - \int_0^1 x^2\, dx = \int_0^1 (\sqrt{x} - x^2)\, dx$$

and finally calculate the integral. There are certainly examples of questions prior to Calculus II where students are required to create an equation instead of the equations being provided, but in Calculus II and later this becomes the rule rather than the exception. This corresponds to the higher levels of S&S, where "general procedures rather than algorithms must be followed" and the questions require “access of relevant knowledge, self-reflection on actions, exploring concepts, and analysis of constraints.”

Junior level classes (prerequisite levels 3, 4, and 5) seemed to have the highest cognitive demands. Once again, this is largely due to the fact that students at this level are responsible for proofs, and also must create or set up problems before solving them. To illustrate why these prerequisite levels have high HOCS ratings and point transformations, consider the following question from prerequisite level 4 (Calculus II) which was rated as a Stein & Smith (S&S) level 4 question:

Evaluate the indefinite integral $\int \frac{t}{\sqrt{2t^2+3}}\, dt$.

First, notice the problem offers no suggestion as to what method to use. The student must realize the u-substitution method is necessary. The student must then let $u = 2t^2 + 3$, which means $du = 4t\, dt$, which in turn means $t\, dt = \frac{1}{4} du$. So the integral turns into $\frac{1}{4}\int \frac{1}{\sqrt{u}}\, du.$

From this point, the student should transform the integral to $\frac{1}{4}\int u^{-\frac{1}{2}}\, du$ in order to use the power rule. The final step is to back substitute for u and add the constant of integration.

$$\frac{1}{4}\int u^{-\frac{1}{2}}\, du = \frac{1}{4}\left[2\sqrt{u}\,\right] + C = \frac{1}{2}\sqrt{2t^2+3} + C$$

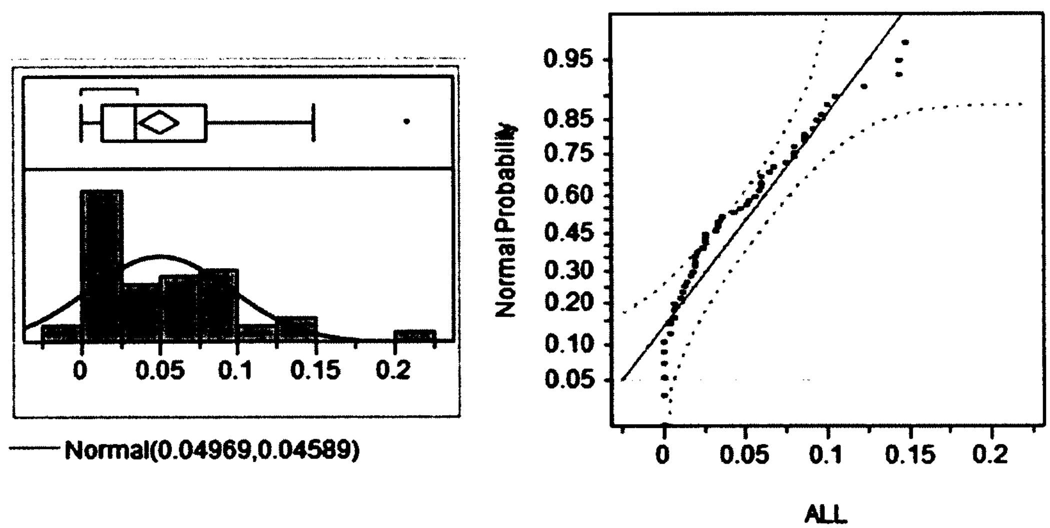
A few of these steps could be considered algorithmic (the power rule, changing a square root to a half power, etc.). However, the student must solve an unpredictable problem by analyzing constraints, accessing relevant knowledge, self-reflecting on actions, and performing some non-algorithmic steps.

There seemed to be a slight decrease in the level of cognitive demand from the junior to senior level classes (prerequisite levels 5 to 6). At prerequisite level 6, the courses seem to emphasize a firm and thorough understanding of definitions and theorems. While these definitions and theorems may be abstract, understanding them is not as conceptually difficult as those in prior prerequisite levels. In Introduction to Analysis I and II, students learn about familiar topics from Calculus I and II, but at a deeper, more rigorous level. Though the courses may be challenging, the students are somewhat familiar with the material. Perhaps another reason for this decrease in cognitive demand is that both of the Analysis courses have students assemble tools from Calculus as the semester progresses. Therefore, it's possible students are more equipped to solve a problem on the final exam than they would be on the midterm.

Analysis of Correlation Between Individual Average HOCS Transformations and Overall Prerequisite Average HOCS Transformations

Pearson's *r* was used to identify any possible correlation between the HOCS classifications using the average point transformations for each HOCS measurement for every exam. A *t*-test was used to find the significance of the correlations, and this test assumes the data are continuous and normally distributed.

Figure 3. Verifying t-test Assumptions



There are some slight violations of normality. There is one outlier at 0.208 (the average point transformation for the CAS classification for prerequisite level 4). There are also a large number of exams with average point transformations between approximately 0.001 and 0.02. However, the diagnostics plot suggests there are no serious violation of normality. Since the t-test is robust against non-normality and there are no serious violations of normality, the results of the test will be valid. Figure 4 shows the correlation between all average HOCS classification transformations, and Figure 5 illustrates the significant correlation between average HOCS transformations.

Figure 4. Correlation Matrix for HOCS Measurements

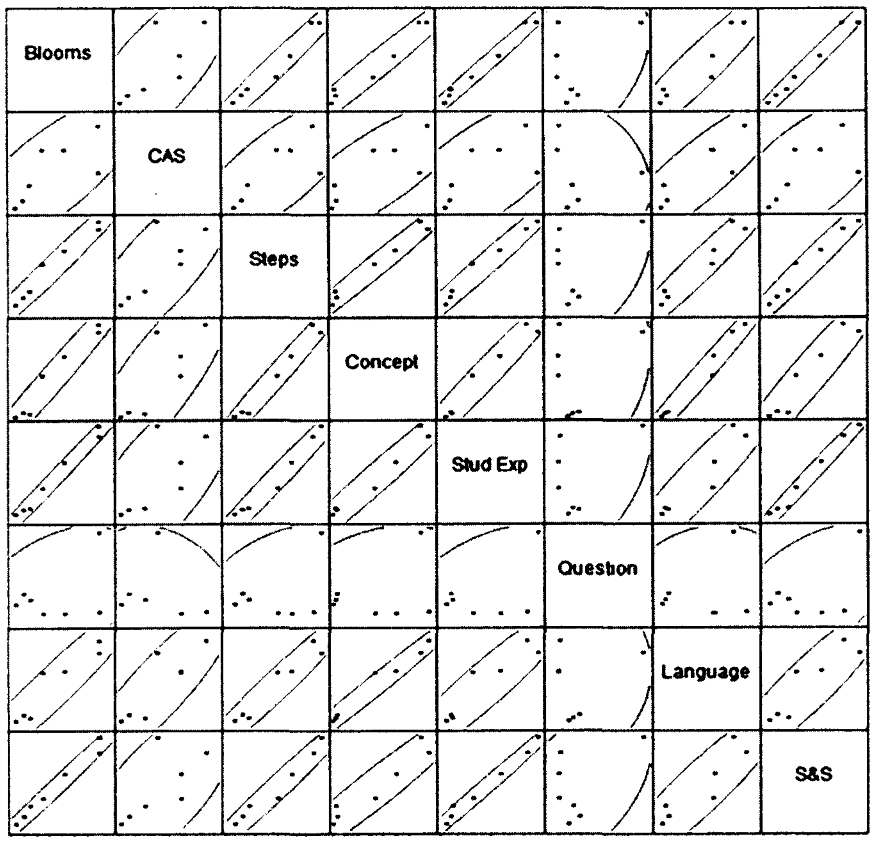


Figure 5. Pearson's Correlation Between HOCS Measurements.

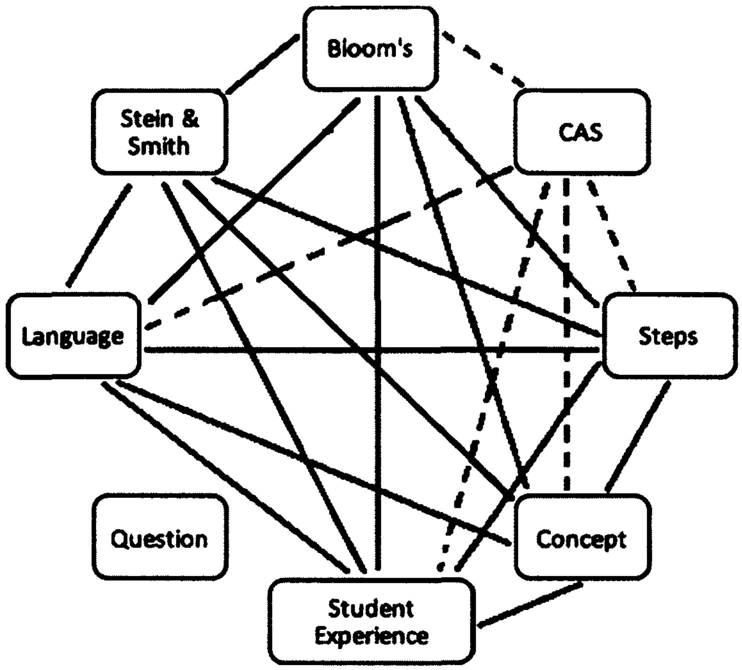


Figure 5. The correlation coefficients, Pearson's *r*, were calculated to find the possible correlations between the measurements of HOCS. A correlation in between 0.7 and 0.9 is denoted by a dotted line, and a correlation greater than 0.9 is denoted by a solid line.

<u>Analysis of Correlation Between Individual Prerequisite Levels and Average HOCS Transformations</u>

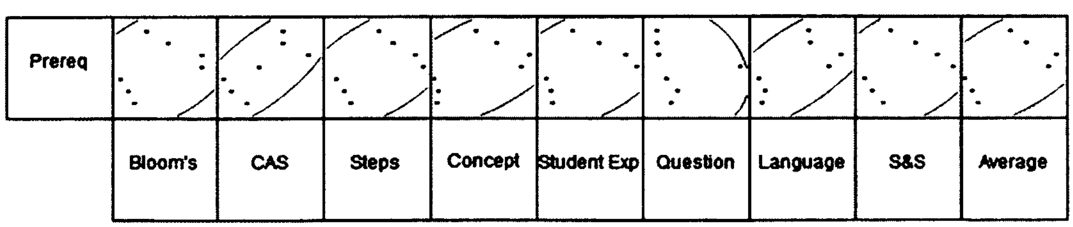
The correlation between prerequisite levels and the average point transformations of each HOCS were also calculated. Since the prerequisite levels are ordinal and the average point transformations are continuous, Spearman's Rho was used to find the correlations (see Table 5).

Table 5

Correlation (Spearman's ρ) Between Prerequisites and HOCS Measurements

HOCS Measurement	$\hat{\rho}$	p-value
Bloom's	0.500	0.253
CAS	0.714	0.071
Steps	0.500	0.253
Concept	0.571	0.180
Student Exp	0.500	0.253
Question	0.739	0.058
Language	0.571	0.180
S&S	0.500	0.253

Figure 6. Spearman's Correlation Between Prerequisites and HOCS Measurements



With a significance level of $\alpha = 0.10$, there was a significant (p-value = 0.07) correlation ($\hat{\rho} = 0.71$) between prerequisites and CAS. In other words, as the number of prerequisites required increases, the ability to solve the questions using the computer software WA decreases.

There was also significant (p-value = 0.06) correlation ($\hat{\rho} = -0.74$) between prerequisites and Question. However, this correlation is largely due to outliers. Some exams for classes with low numbers of prerequisites (specifically College Algebra, Plane Analytic Geometry, Pre-Calculus, Calculus I, Calculus II, and Introduction to Modern Math) had a few questions that were possibly confusing or somewhat open to interpretation (Question level 2). One exam (Calculus I) had one question rated as Question 3 (very possibly confusing or open to interpretation). This question was stated as:

If $\int_1^{10} f(x) = 5\,dx$ and $\int_1^{3} f(x) = -2\,dx$, what is $\int_1^{10} f(x)\,dx$?

There is one small typo; the differential dx is in the wrong location twice. The question should be:

If $\int_1^{10} f(x)\,dx = 5$ and $\int_1^{3} f(x)\,dx = -2$,

what is $\int_1^{10} f(x)\,dx$?

This is a small mistake, and the students most likely understood what was being asked. However, this is technically Question 3 since there is a possibility this mistake could lead to confusion (unintended, invalid difficulty).

The vast majority of questions were rated as Question level 1 (not open to interpretation, a clearly stated question). So the negative correlation is artificial in that as the number of prerequisites increases, the questions do not become more clearly worded and less open to interpretation. This negative correlation is only a product of those few outliers.

There was not a significant (p = 0.22) correlation ($\hat{\rho} = 0.54$) between prerequisite levels and the average HOCS ratings. In other words, courses with more required prerequisites are not necessarily associated with higher HOCS demands.

Model Refinement

In order to streamline future analyses, it would be useful to reduce the number of HOCS classifications while retaining the maximum amount of information possible. This is a challenge because using too few measurements would result in too little explanatory power. Using too many measurements is overly time consuming and can obscure patterns in the data. The R^2 value (which has a range of [0,1]) was maximized in order to identify the most explanatory subsets of HOCS classifications using regression and multiple regression. The HOCS measurements will be used to predict the average HOCS measurements (without the Question rating) for each prerequisite level.

<u>Single Most Explanatory HOCS Measurement</u>

In order to find the single most explanatory HOCS measurement, linear regression was used to model the relationship between the average HOCS transformation values at each prerequisite level (the predictor variables) and the overall average of the HOCS measurements at each prerequisite level (the response variables) (with the Question rating not included in the overall average). Note that since the Question measurement was of such little comparative value (in terms of explaining variation and predicting future measurements), it was not considered in this analysis.

Table 6

Single Most Explanatory HOCS Measurements

Rank	Measurement	R^2
1	Concept	0.991
2	Steps	0.982

3	Language	0.974
4	Bloom's	0.962
5	Student Exp	0.947
6	S&S	0.926
7	CAS	0.704

Most Explanatory Subsets of HOCS Measurements

Finding the most explanatory subset of HOCS measurements is more difficult than finding the single most explanatory HOCS measurement. A common misconception to obtain the most explanatory subset of *n* measurements would be to simply take the *n* most explanatory single measurements. This method is incorrect because the two measurements may be highly correlated (the two most explanatory HOCS measurements Concept and Language are in fact highly correlated with r = 0.987), and so including the second most explanatory HOCS measurement may provide little extra value. Two of the most popular methods of model refinement (forward stepwise and backward stepwise) are used. In the forward stepwise method, the single most explanatory classification is chosen. Then all other classifications are included individually until the one producing the highest R^2 value is found, and so on. In the backwards elimination method, all classifications are chosen. Then each classification is removed one at a time to see which (if any) would detract the least from the predictive power.

Concept and Steps Required were removed before searching for the best subsets since they were both highly subjective categories and difficult for the researcher to obtain. Only the five best subsets of each size are reported (Table 6). In summary, the five classification systems considered were

- Bloom's Modified Taxonomy
- Computer Algebra System (CAS)
- Language
- Stein and Smith (S&S)
- Student Experience

Table 7

Most Explanatory Subsets of HOCS Measurements

Size of Subset	R^2	Measurements Used
2	0.999	CAS, Stud Experience
2	0.996	Bloom's, Language
2	0.996	S&S, Language

2	0.992	Student Experience, Language
2	0.991	Bloom's, CAS
3	1.0	CAS, S&S, Language
3	0.999	CAS, Stud Experience, Language
3	0.999	Bloom's, CAS, Stud Experience
3	0.999	CAS, Stud Experience, S&S
3	0.998	Bloom's, CAS, Language
4	1.0	Exclude Bloom's
4	1.0	Exclude Stud Experience
4	0.999	Exclude S&S
4	0.999	Exclude Language
4	0.998	Exclude CAS
5	1.0	All

It may seem strange that only two HOCS classifications can encapsulate virtually all of the information contained in the entire list of eight HOCS measurements. In particular, why isn't Bloom's Modified Taxonomy needed at all in some cases? The answer is that Bloom's is highly correlated with the measurements that are included; CAS (r = 0.739), Student Experience (r = 0.992), Language (r = 0.944), and S&S (r = 0.990). Notice that while CAS is the least explanatory measurement alone, it's included in all of the most explanatory subsets. This is due to the fact that CAS isn't highly correlated ($r \geq 0.9$) with any of the other classifications. Among the five HOCS measurements considered, Student Experience was the most difficult for the researcher to obtain (due to the subjectivity of the classification), so subsets including this classification were avoided.

Most Explanatory Overall Subsets

In summary, the following subsets of HOCS classifications seemed to be the best in terms of explaining the variation in the data as well as being the easiest measurements to collect by the researcher.

Table 8

Most Explanatory Overall Subsets

Size of Subset	R^2	Measurements Used
1	0.975	Language
2	0.996	Bloom's, Language
2	0.996	Language, S&S
2	0.991	Bloom's, CAS
3	1.0	CAS, Language, S&S
3	0.998	CAS, Language, Bloom's
4	1.0	Bloom's, CAS, Language, S&S

From this point, the choice for the best subset becomes a personal preference. In order to illustrate the predictive ability of these subsets, consider the model for the three classification subset of CAS, Language, and S&S ($R^2 = 1.0$):

$$0.12899(CAS) + 0.36946(Language) + 0.45931(S\&S) + 0.00257 \quad (2)$$

Using this model to predict the average HOCS point transformation for prerequisite level 4 yields 0.09038 when the actual value was 0.08577 (a difference of 0.00461).

Since there is a linear relationship between the actual average HOCS transformations and the predicted average HOCS transformations (Figure 7), Equation 2 appears to be a good model. Since the residuals for the predicted values (Figure 8) are small, symmetrically distributed about 0, and follow no clear pattern, there is further evidence Equation 2 is a good model.

Figure 7. Actual vs. Predicted HOCS Transformations

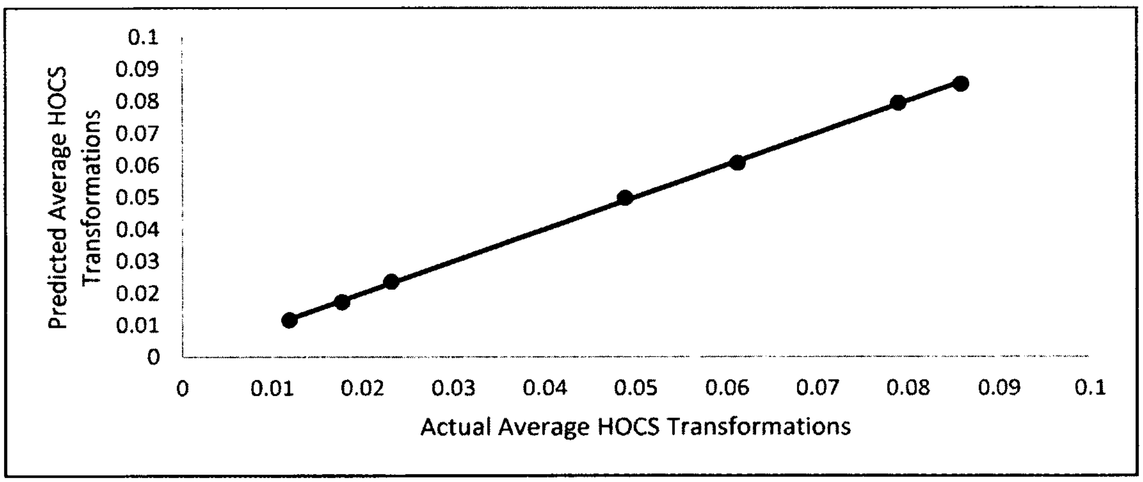
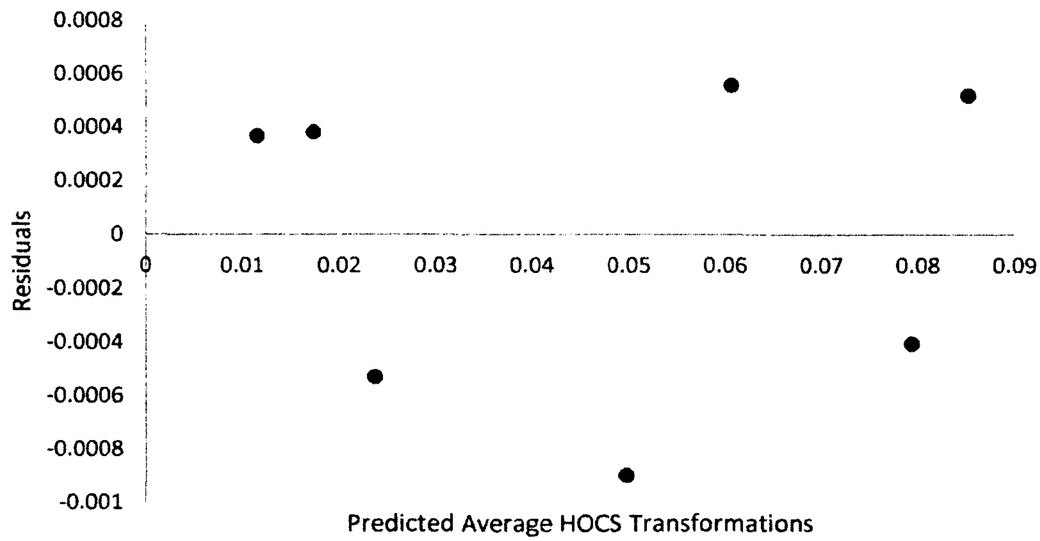


Figure 8. Predicted Average HOCS Transformations vs Residuals



<u>Limitations of the HOCS Classifications and Exam Analysis</u>

Some of the HOCS classifications were by necessity subjective. For example, the Student Experience ratings were decided based on the researcher's memory of the courses. Furthermore, a question that would normally be rated as Student Experience level 2 or 3 may have been covered in detail by the professor in a review session prior to the final exam. This fact could possibly lower the Student Experience rating from a level 3 to 2, or level 2 to 1. The researcher attempted to remove some of this subjectivity by removing some of the most subjective HOCS classifications from the most explanatory subsets. Other solutions to the problem of subjectivity could be to communicate with the professors, or the questions could be classified by a group instead of an individual researcher.

It is possible some of the final exams did not accurately represent the HOCS demand of the overall course. This final exam analysis is only an indicator (but not a true measurement) of the HOCS demand of the courses.

CHAPTER 5

Discussion and Final Remarks

There seemed to be some common themes in questions requiring higher levels of cognition. First, these questions requiring HOCS often referred to "real-life" scenarios. For example, notice the difference between the following two versions of the same question:

1. Find the y value when the derivative of $y = -x^2 + 13x + 5$ is 0.
2. The height of a ball thrown into the air is given by the equation $y = -x^2 + 13x + 5$. What is the maximum height the ball will achieve?

The solution process for both of these questions is exactly the same. However in the first question, the student is given strong hints on how to proceed (find the derivative, set it equal to zero, and proceed). In the second question, the onus is on the student to know a derivative is needed. Simply putting the same problem in context of a "real-life" word problem has increase the level of HOCS demand.

Secondly, one of the most challenging cognitive hurdles seems to be formulating mathematical expressions from English descriptions. Consider the following question from Statistical Modeling:

Fisherman working certain parts of the Atlantic Ocean sometimes come into contact with whales. Ideally, they would like to scare off whales, but not fish they are trying to catch. One strategy to do this is to transmit underwater the sound of a killer whale. Of the 52 instances of doing this, the technique worked 24 times. However, experience has shown in the past that 40% of whales sighted near fishing boats will leave of their own accord. Test to see if the underwater transmission strategy increases the chance of scaring off whales.

From only this English description of a problem, the student must create a null hypothesis, an alternative hypothesis, a test statistic, etc. The question offers no suggestion as to what route to take or formulas to use. The onus is on the student to translate from the English description to create a statistical test, conduct the test, and finally interpret the test.

Lastly, some of the most cognitively challenging problems combine many different mathematical ideas into a single cohesive question. For example, in Differential Equations, students are asked to solve a system of equations where some of the equations contain derivatives. In order to solve this type of problem, the student must have mastered the ideas of all of these individual topics in order to navigate them all at once.

The research of this paper supports the following recommendations. Lemons states “one of the best ways to help students develop HOCS is to make HOCS questioning a regular part of (the students’) coursework” (p. 47). One of the conclusions from the study of California universities concluded “we need to disseminate information on teaching for critical thinking within particular disciplines (such as math)” (Paul et al.).

Another possible way to improve the thinking skills of students would be to increase the amount of dialogue among faculty on the subject. Professors most likely discuss teaching methods for classes they teach more than the classes they don’t. It’s possible there is little dialogue between professors with vastly different educational philosophies who could learn from one another.

For future analyses, it would be more efficient to use a subset of the HOCS classifications. This subset of HOCS classifications could be used on more facets of a course than just the final exam.

A New HOCS Classification

When this exam analysis was almost complete, another HOCS classification was found. It is called the Rigor/Relevance

Framework and was created by the International Center for Leadership in Education in 1996 "to guide schools on how to deliver a curriculum that is both rigorous and relevant" (Daggett, 2011, p. 1). The results of this research agree with the Rigor/Relevance Framework. In particular, higher HOCS demands are associated with both higher levels of Bloom's Taxonomy as well as applications to "real-world unpredictable situations" (Daggett, 2011, p. 2). The lowest level (Quadrant A) of this framework "focuses on rules, control, teaching/teachers, compliance, and input" whereas the highest level (Quadrant D) "focuses on results, empowerment, learning/students, engagement, and output" (Dagget, 2011, p. 3).

REFERENCES

Arum R., Roksa J. (2011). Academically Adrift. *Limited Learning on College Campuses*, Chicago: University of Chicago Press.

Anderson et al. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn and Bacon. ISBN 978-0-8013-1903-7.

Belkin, D. (2015, January 16). Test Finds College Graduates Lack Skills for White-Collar Jobs. *The Wall Street Journal.* Retrieved from http://www.wsj.com/articles/test-finds-many-students-ill-prepared-to-enter-work-force-1421432744

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.

Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. London: School of Education, King's College London.

Black, P., Harrison, C., Lee, C., Marshall B., and Wiliam, D. (2002). "Working Inside the Black Box: Assessment for Learning in the Classroom," *Phi Delta Kappan*, Vol. 86, No. 1, September 2004, pp. 9-21.

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

Cardellichio, T., & Field, W. (1997). Seven Strategies That Encourage Neural Branching. *How Children Learn, 54*(6), 33-36.

Connor-Greene, P. (2000). Assessing and Promoting Student Learning: Blurring the Line Between Teaching and Testing. *Teaching of Psychology*, 27(2), 84-88.

Cotton, Kathleen (1988). Classroom Questioning. www.learner.org/workshops/socialstudies/pdf/session6/6.ClassroomQuestioning.pdf

Daggett, W. (2011). Living in Quadrant D. *International Center for Leadership in Education*.

Elliot and Stimpson. (2011). Retreived from http://support.battlecreek.k12.mi.us/pluginfile.php/2985/mod_page/content/1/BestPractices/Math/cognitive_demand.pdf

Fellenz, M. (2004). Using assessment to support higher level learning: The multiple choice item development assignment. *Assessment & Evaluation in Higher Education*, 29(6), 703-719.

Fisher-Hoch, H., & Hughes, S. (1996). What Makes Mathematics Exam Questions Difficult. *BERA: British Education Research Association*.

Framework for 21st Century Learning. (2009). Partnership for 21st Century learning (P21). Retrieved from http://www.p21.org/about-us/p21-framework

Hannel, I. (2009). Insufficient Questioning. *Phi Delta Kappan*, 91(3), 65-69.

Hastings, S. (2003). Questioning. *The Times Educational Supplement*.

International Society for Technology in Education. (2002). *National Educational Technology Standards for Teachers: Preparing teachers to use technology*. Eugene, OR: ISTE.

Karaali, G. (2011). An Evaluative Calculus Project: Applying Bloom's Taxonomy to the Calculus Classroom. *PRIMUS*, 21(8), 719-731.

Kokol-Voljc, V. (1999). Exam questions when using CAS for school mathematics teaching.

Krathwohl, David R. (2002). "A revision of Bloom's taxonomy: An overview". *Theory Into Practice* (Routledge) 41 (4): 212–218. doi:10.1207/s15430421tip4104_2. ISSN 0040-5841.

Lemons, P., & Lemons, J. (2013). Questions for Assessing Higher-Order Cognitive Skills: It's Not Just Bloom's. *Cell Biology Education*, 12, 47-58.

National Association of Colleges and Employers (NACE). (2015, April). *Job Outlook 2015 Spring Update: Hiring to Increase 9.6 Percent*. Bethlehem, PA: Gray, K.

Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions?. *BMC Medical Education*, 49.

Paul, R., & Binker, A. (1990). Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World. Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R., & Elder, L. (2008). The Miniature Guide to Critical Thinking Concepts and Tools. *Foundation for Critical Thinking Press.*

Paul, R., Elder, L., & Bartell, T. (1997). California Teacher Preparation for Instruction in Critical Thinking: Research Findings and Policy Recommendations California Commission on Teacher Credentialing, Sacramento California, 1997. http://www.criticalthinking.org/pages/center-for-critical thinking/401

Scriven, M., & Paul, R. (1987). 8th Annual International Conference on Critical Thinking and Education Reform. Retrieved December 8, 2015, from http://www.criticalthinking.org/pages/defining-critical-thinking/766

Stein, M., & Smith, M. (2000). Implementing Standards-Based Mathematics Instruction: A Casebook for Professional Development (Second ed.). Teachers College Press.

Sousa, D. (2011). How the Brain Learns (4th ed., p. 266). Thousand Oaks, California: Corwin.

VITA

Chance Bradford received his Associate of Science degree from Northeast Texas Community College in 2010. He then attended the University of Texas at Austin for one year. He received his Bachelor of Science degree from Stephen F. Austin State University in May 2013 with a major in Mathematics and a minor in Secondary Education. In the fall of 2013, he entered the Graduate School of Stephen F. Austin State University and plans to receive the degree of Master of Science in December 2015.

Permanent Address: 4203 NE Stallings Dr Apt 302

Nacogdoches, TX 75965-1642

APA Style

This thesis was typed by Chance Bradford