8-2014

An Inventory of Student Recollections of Their Past Misconceptions as a Tool for Improved Classroom Astronomy Instruction

Andrej Favia

Recommended Citation

Favia, Andrej, "An Inventory of Student Recollections of Their Past Misconceptions as a Tool for Improved Classroom Astronomy Instruction" (2014). *Electronic Theses and Dissertations*. 2200.
http://digitalcommons.library.umaine.edu/etd/2200

AN INVENTORY OF STUDENT RECOLLECTIONS OF THEIR PAST MISCONCEPTIONS AS A TOOL FOR IMPROVED CLASSROOM ASTRONOMY INSTRUCTION

By
Andrej Favia
B.A. University of Southern Maine, 2007

A DISSERTATION
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
(in Physics)

The Graduate School
The University of Maine
August 2014

Advisory Committee:

Neil F. Comins, Professor of Physics and Astronomy, Co-Advisor

Geoffrey L. Thorpe, Professor Emeritus of Psychology, Co-Advisor

David J. Batuski, Professor of Physics and Astronomy

James P. McClymer, Associate Professor of Physics and Astronomy

Andrew A. West, Assistant Professor of Physics and Astronomy

DISSERTATION ACCEPTANCE STATEMENT

On behalf of the Graduate Committee for Andrej Favia, I affirm that this manuscript is the final and accepted dissertation. Signatures of all committee members are on file with the Graduate School at the University of Maine, 42 Stodder Hall, Orono, Maine.

Dr. Neil F. Comins, Professor, Physics and Astronomy Date

Dr. Geoffrey L. Thorpe, Professor Emeritus, Psychology Date

LIBRARY RIGHTS STATEMENT

In presenting this dissertation in partial fulfillment of the requirements for an advanced degree at The University of Maine, I agree that the Library shall make it freely available for inspection. I further agree that permission for “fair use” copying of this dissertation for scholarly purposes may be granted by the Librarian. It is understood that any copying or publication of this dissertation for financial gain shall not be allowed without my written permission.

Signature: Andrej Favia

Date: August 6, 2014

AN INVENTORY OF STUDENT RECOLLECTIONS OF THEIR PAST MISCONCEPTIONS AS A TOOL FOR IMPROVED CLASSROOM ASTRONOMY INSTRUCTION

By Andrej Favia

Dissertation Co-Advisors: Dr. Neil F. Comins
Dr. Geoffrey L. Thorpe

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Physics)
August 2014

My Ph.D. research is about examining the persistence of 215 common misconceptions in astronomy. Each misconception is based on an often commonly-held incorrect belief by college students taking introductory astronomy. At the University of Maine, the course is taught in alternating semesters by Prof. Neil F. Comins and Prof. David J. Batuski.

In this dissertation, I examine the persistence of common astronomy misconceptions by the administration of a retrospective survey. The survey is a new instrument in that it permits the student to indicate either endorsement or rejection of each misconception at various stages in the student's life. I analyze data from a total of 639 students over six semesters. I compare the survey data to the results of exams taken by the students and additional instruments that assess students' misconceptions prior to instruction. I show that the consistency of the students' recollection of their own misconceptions is on par with the consistency of responses between prelims and the final exam. I also find that students who report higher increased childhood interest in astronomy are more likely to have accurate recalls of their own past recollections.

I then discuss the use of principal components analysis as a technique for describing the extent to which misconceptions are correlated with each other. The analysis yields logical groupings of subtopics from which to teach. I then present a brief overview of item response theory, the methodology of which calculates relative difficulties of the items. My analysis reveals orders to teach the associated topics in ways that are most effective at dispelling misconceptions during instruction. I also find that the best order to teach the associated concepts is often different for high school and college level courses.

ACKNOWLEDGEMENTS

Above all, I am grateful for Campus Crusade for Christ, the Navigators, and Alpha, for bringing me up spiritually in 2013 from a low place in my life and turning to Christ; I haven’t looked back since. I would like to acknowledge my committee: Neil F. Comins, for his attention to detail and commitment to understanding misconceptions, Geoffrey L. Thorpe, for his expertise in statistical analysis, David J. Batuski, for volunteering countless hours of his time to prepare special video lectures, and equipping me with the resources necessary to teach AST 216 (Spring 2014), James P. McClymer, for setting clear goals for my time as a graduate student, and Andrew A. West, currently at Boston University, for guiding me on analyzing hypervelocity M dwarf candidates. I acknowledge Philip M. Sadler, my external dissertation reader, for supporting my research, and David Hiebeler for his expertise in **R**. I acknowledge members of the Physics Education Research Laboratory for critiquing my research methodology, and to Jessica W. Clark for contributing her time to the inter-rater analysis component of my dissertation. I am also grateful to have worked with David Sturm, for his many encouraging conversations on my research, his experience in teaching and managing the labs, and for allowing me to work with him during numerous enjoyable physics shows over the years. Lastly, I thank my immediate family for allowing me to visit when I just need to relax or do something recreational in the White Mountains.

TABLE OF CONTENTS

Chapter

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NFC	Prof. Neil F. Comins	16
DJB	Prof. David J. Batuski	16
sA	statement in astronomy	22
ABI	Astronomy Beliefs Inventory	23
MMRS	mean misconception retainment score	31
ANOVA	ANalysis Of VAriance	32
CIAS	Childhood Interest in Astronomy Survey	60
GLT	Prof. Geoff L. Thorpe	64
BHGP	Black Holes and Galaxies Pretest	67
PCA	principal components analysis	119
PA	parallel analysis	121
IRT	item response theory	221
1PL	one-parameter logistic	224
ICC	item characteristic curve	224
2PL	two-parameter logistic	226
CRC	characteristic response curve	227
GRM	graded response model	227
MGC	misconception group coordinate	240
MANOVA	Multivariate ANalysis Of VAriance	331

CHAPTER 1

INTRODUCTION

1.1 Purpose of the research

The purpose of this research is to study the persistence of misconceptions that students bring to the college astronomy classroom, with the focus of my research on students enrolled in the introductory astronomy course, at the University of Maine. The core of this research is the development and implementation of a comprehensive inventory of misconceptions in astronomy. The goal of the study is to analyze the inventory responses to determine an optimal way to present topics in astronomy that dispels the misconceptions most effectively. My research project involves an in-depth analysis of the persistence of misconceptions held by these students in various topics in astronomy, such as stars, the solar system, the Moon, the Earth, other planets in our solar system, the Sun, galaxies, and black holes. The contribution of this research to the field of astronomy education is to inform astronomy instructors on the nature of students' misconceptions, so that instructors may know how to target misconceptions in astronomy more effectively.

1.2 Background

Since the 1980s, the process of how students learn has been studied in several topics within the field of physics and astronomy education research. Such topics include perceptions of motion (diSessa, 1982; White, 1982), the colors of objects (Anderson & Smith, 1988), and heat (Kempton, 1986), among other topics in astronomy as studied by Flavell, Green, and Flavell (1986), Posner, Strike, and Hewson (1982), Sadler (1998), Vosniadou and Brewer (1992), and Vosniadou (1994). The subjects of the studies range from children to adults. These studies draw two common conclusions about the learning process. First, learning is a complex process that has no “one size fits all” rule on how to teach the relevant material to the class. Second, even after instruction, students in the class may retain any one of a multitude of inappropriate models to explain their observations of relatively simple physical phenomena. These research studies suggest that in order to teach effectively, teachers should examine the learning processes of their students more carefully. In this Section, I introduce definitions of terms used in the science education literature to describe how students construct their models.

1.2.1 Preconceptions

In a typical introductory course in astronomy, students are exposed to a variety of factoids and concepts related to astronomy, with some concepts (e.g., those regarding black holes) being more obscure than others. A *factoid* is a single bit of factual information. Some factoids may be relatively obvious (e.g., the Earth has only one Moon), while others may be relatively obscure (e.g., black holes lose mass due to Hawking radiation) and may require that students learn extra concepts for the factoids to seem meaningful. A *concept* is a mental picture of one or more physical objects (e.g., a light bulb) or abstract qualities (e.g., heat). Because concepts are mental pictures, they serve as ideas that are introduced to explain scientific phenomena (Eylon & Linn, 1988, p. 252). As discussed by Driver (1983, p. 66) and J. P. Prather (1985, p. 11), a *preconception* is an idea or opinion about the cause or effect of a natural phenomenon. A simple example is the expectation that more force leads to more motion. According to J. P. Prather and references therein, students often develop preconceptions about scientific phenomena “through their own observational and intellectual prowess without formal instruction in science” (p. 24). Outside the classroom, students may acquire both accurate information and misinformation from other sources, such as life experiences, and various media presentations (Comins, 2001; Libarkin, Asghar, Crockett, & Sadler, 2011). That students develop their own preconceptions without guided instruction suggests that students may set their own initial expectations of scientific phenomena early in their lives.

Preconceptions are limited by the extent of one's own mental representation of the phenomenon. A more complete mental representation requires synthesizing new knowledge with pre-existing knowledge, or modifying existing knowledge to fit observations. In education research, the instructor is primarily concerned with how the student observes the world and acquires knowledge from it in order to learn, a process termed *conceptual change*. In both physics and astronomy, the primary goal of studying conceptual change is to capture how students acquire knowledge, and how students use the knowledge that they have acquired, either previously or during live instruction, to explain physical phenomena that they observe. In the learning process, one often builds a *model*, or *mental model*, as more formally stated in the literature (Vosniadou, 1994), which is an explanation of how something works that is supported by observational evidence.

Conceptual change deals with the process of refining or replacing one's mental models. The structure underlying the construction of mental models is described by Vosniadou (1994) and summarized below, with four key elements influencing the process of conceptual change:

1. Observations about the nature of being or how the world exists are called *presuppositions*. Presuppositions can be divided into two categories: ontological presuppositions (e.g., "up" vs. "down"), and epistemological presuppositions (e.g., "things are as they appear to be"). The set of all such presuppositions composes one's *framework theory*.

2. The set of all scientific observations about the world (e.g., "objects can feel hot, warm, or cold," or "the sky is located above the ground") compose one's *specific theory*.

3. A *belief* is an unsubstantiated claim; it is an assertion of the truth based merely on acceptance or faith, rather than on supporting evidence. Beliefs develop when one attempts to reconcile one's own framework theory with one's own specific theory, that is, a collection of beliefs that one has that "describe the properties and behavior of physical objects" (p. 47). Beliefs may be scientifically valid, though as Vosniadou acknowledges in her study of beliefs endorsed by elementary school children, many students' beliefs are incorrect. For example, in the context of heat, some students believe that "hotness" and "coldness" are two different properties of objects (p. 48). In the context of the shape of the Earth, many students believe that the Earth is flat.

4. One finally synthesizes a mental model when one has enough supporting evidence to build an explanation based on one's own belief. Example mental models include a "two agents" model for heat transfer (i.e. with hot and cold as separate entities) and a model of the Sun physically moving up and down in space to explain the day-night cycle. Conceptual change is the result of modifying one's own mental model.

The study by Vosniadou shows that the structure of the learning process uses these same four elements (framework theory, specific theory, beliefs, mental models) for describing, e.g., the day-night cycle, the concept of heat, and the concept of force. Her study suggests that instructors can analyze mental models constructed by students throughout many topics within the fields of physics and astronomy.

Vosniadou acknowledges that in the model-building process, students may attempt to reconcile their own preconceptions with new knowledge in such a way that students develop their own incorrect explanations. In multiple studies, Vosniadou et al. have shown that children develop and retain the misinformation and, from it, form one of many possible "synthetic models," e.g., regarding the Earth-Sun system (Vosniadou, 1994) and the formation of stars (Vosniadou, Vamvakoussi, & Skopeliti, 2008). For example, in an investigation of the beliefs endorsed by elementary school children, Vosniadou (1994) report that some students believe that (i) the Sun actually goes up and down on the sky and (ii) the Earth rotates, and that these mixed true-false beliefs lead to a synthetic model, e.g., that “the sun and the moon move up/down to the other side of the earth” (p. 169). An example of the consequences of developing an incorrect explanation is given by Eaton, Anderson, and Smith (1983), who conducted a study on fifth graders receiving instruction on light and photosynthesis. As noted by J. P. Prather, most students in the study “never made the connection between light and food making; they had explained the phenomenon to their own satisfaction without it” (pp. 10-11). Interestingly, J. P. Prather notes that these students are some of many who develop *naïve* beliefs to explain scientific phenomena, which is to say, the students feel like they are explaining the truth but don’t know that they aren’t (p. 4).

Because many preconceptions develop prior to formal instruction, students may bring their own preconceptions to high school and college classrooms, and perhaps retain them even after instruction. A number of studies, including those by Bailey, Prather, Johnson, and Slater (2009), Sadler (1998), Sadler et al. (2010), and Wallace, Prather, and Duncan (2011) have shown that any number of inappropriate preconceptions in astronomy may persist after instruction, and some of these preconceptions may persist all the way to college. The study by Bailey et al. assessed pre-instructional ideas about stars and star formation held by 2,200 non-science majors taking an introductory astronomy course (“pre-instructional” meaning prior to starting the course). Bailey et al. observed that students often bring misinformation to the classroom (e.g., a star is a “burning ball of gas”). As another example, the study by Sadler (1998) tested 1,250 students ranging from 8th through 12th grade on their beliefs in astronomical phenomena, such as the cause of the seasons and night and day. Sadler found that students tend to harbor one of many common naïve beliefs to explain a single phenomenon; however, some students may abandon an incorrect belief in favor of yet another incorrect belief. Because both incorrect beliefs are based on preconceptions, Sadler terms them *alternate conceptions*.

1.2.2 Misconceptions

Interestingly, Sadler uses the term “misconceptions” to refer to “conceptions that are quite different from their teachers’ or those of scientists and that these ideas are resistant to change, even in the face of the most dogged efforts” (p. 265). While they do not define the term, “misconceptions,” as Smith III, diSessa, and Roschelle (1993) write, are sometimes discussed in the context of similar terms, such as preconceptions, alternative conceptions, and naïve beliefs. The fundamental differences in these terms are in how researchers “have characterized the cognitive properties of student ideas and their relation to expert concepts” (p. 119), where cognitive refers to how one makes meaning from information. Eylon and Linn (1988) note that a variety of terms, including misconceptions, alternative conceptions, naïve conceptions, and alternative frameworks, have been used in previous research to describe “the qualitative differences among the concepts students use to explain scientific phenomena” (p. 252).

In the previous literature review, researchers tend to interpret misconceptions from the standpoint of how they think students learn. For example, the framework described by Vosniadou (1994) suggests that learning involves the strategic refinement of one’s own mental models, and that misconceptions are attempts to reconcile scientific information with false information contained within one’s own framework theory. An alternative framework known as constructivism, discussed in Section 1.2.4, suggests that learning “involves the interpretation of phenomena, situations, and events...through the perspective of the learner’s existing knowledge” (Smith III et al., 1993, p. 116). Smith III et al. note further that learning “involves the acquisition of expert concepts and the dispelling of misconceptions” (p. 122). Other learning frameworks, including that described by Strike and Posner (1992), suggest that conceptual change requires that the learner become dissatisfied with current concepts, and that learning occurs when a new conception seems to solve the current problem and encourages new ways of looking at natural phenomena.

In common with all of these conceptual frameworks is the notion that misconceptions are *deep seated* incorrect beliefs. That is to say, they persist so much with the students that students may retain inappropriate expectations of natural phenomena even after instruction. In their study, Smith III et al. (1993) observed that misconceptions persist even after instruction, and that misconceptions “have a strong influence on how student learning is currently evaluated.” That is to say, misconceptions interfere with learning. This interpretation is also supported by the results of the previous studies (Eylon & Linn, 1988; Sadler, 1998; Bailey et al., 2009). Sadler also demonstrated that some misconceptions tend to persist more so than others. Briefly, Sadler (1998) showed that misconceptions related to the day-night cycle are easier to unlearn than misconceptions related to the seasons. Other studies, including those by LoPresto and Murrell (2011), have shown that incorrect naïve beliefs (including those which reject factoids such as “stars eventually die”) also span a broad range of persistence, and that some incorrect beliefs tend to persist just as much as misconceptions. These results are consistent with the discussion by Sadler, who noted that some factoids are “connected to mental models” and that “such facts should not be taught in isolation, but only to help move students to more powerful and accurate models” (p. 286).

Complicating matters worse is that multiple researchers have alternate, often loosely-worded, definitions of the term “misconception.” For example, Comins Comins (2001) presents a list of definitions for the term “misconception,” the list of which he developed after asking researchers at a conference on misconceptions in science and mathematics in 1994. Comins presents a list of thirteen such definitions (p.

55), which include “any belief that is untrue,” “a mistaken idea,” “an incorrect mental construct,” “a misunderstanding,” “only incorrect beliefs that are deep-seated in our minds,” “a conception or belief that produces a systematic pattern of errors,” and others. Additional definitions of the term “misconception” have since been documented (Comins, 2014). Formally, Comins defines a misconception as “any deeply held belief that is inconsistent with currently accepted scientific concepts. Deeply held beliefs are distinctly different from superficial ones, such as details we might memorize for an exam and then promptly forget” (p. 56). This definition is consistent with the premise that misconceptions are *deep seated* and avoids grounding the term in a particular learning framework.

Given all of these considerations, it is little wonder why the term “misconception” is hard to define. For the purposes of my research, I intend to examine a very broad collection of common incorrect beliefs in astronomy. Some of these beliefs are simply naïve, while others are the result of models used to describe astronomical phenomena. While misconceptions may be defined in terms of the learning framework, multiple unique learning framework theories have already been developed, each with a different basis from which misconceptions are defined. Hence, rather than speak about misconceptions from the associated learning framework, I intend to examine the collection of incorrect beliefs for their tendency to persist with students. *For the purpose of establishing a working definition of a misconception in astronomy, throughout my dissertation I have elected to refer to the definition by Comins (2001, p. 56).*

1.2.3 Unlearning misconceptions vs. learning the correct science

While misconceptions have varying degrees of persistence, the processes of unlearning (or dispelling) the misconceptions, which is to abandon them after being made aware of their inadequacies, and learning the correct science, are not necessarily one and the same. As Clement (1982) notes, in cases where preconceptions interfere with learning, learning “becomes a process in which new concepts must displace or be remolded from stable concepts that the student has constructed over many years” (p. 66). As J. P. Prather (1985) adds, “the first step in the process, [Clement] declared, is helping students articulate their preconceptions so they may become *conscious* of the inadequacies of any misconceptions they may hold. Then, he implied, the correct alternatives, when presented through conventional instructional means, may be perceived as meaningful new concepts to be learned” (pp. 14-15).

The preceding passage by J. P. Prather emphasizes the difference between unlearning of incorrect information and learning the correct science. The unlearning component consists of *drawing students’ attention to inadequacies as they may pertain to their model*. An example of how one would accomplish this for a student who believes that “all stars last forever” would be to assume that the student’s expectation is true, then illustrate observational evidence of supernovae as violent explosions of what used to be extremely massive stars. As Smith III et al. (1993) reminds us, it is important that instructors not overlook this step, because misconceptions interfere with learning. In a more recent study, Sadler et al. (2010) conducted a study of 7599 students and their 88 teachers spanning grades 5-12 on a variety of astronomy topics. The researchers refer to a number of educational standards, notably the NRC Standards and AAAS Benchmarks, in assessing the required knowledge for “astronomical literacy.” Based on these standards and prior knowledge of the misinformation brought by students to the classroom, the researchers designed a series of multiple-choice tests with incorrect answers that directly target misinformation held by students. These targeted incorrect answers are called *distractors*. The study showed that “teachers generally overestimate the performance of their students” (p. 17), in that teachers fail to recognize the extent of their students’ misconceptions. Their findings suggest that *teachers need to be made more aware of their students' misconceptions in teaching astronomy topics*.

J. P. Prather (1985) continues on pp. 25-26 to outline a method for learning the correct science. Briefly, instructors should deploy "teaching strategies that would help students discover and articulate their preconceptions of science," which is to say, teachers should encourage students to communicate their ideas. A few methods, which I discuss in the next paragraph, exist to engage students in the discovery and articulation processes. J. P. Prather continues by noting that students should examine the strengths and weaknesses of their own ideas and those of scientists. Both J. P. Prather and Eaton et al. (1983) encourage the use of astronomy laboratory exercises as learning activities; J. P. Prather specifically encourages instructors "to conduct laboratory exercises comparing the predictive efficacy of [the students'] preconceptions with the conventional science, thereby learning for themselves the superiority of the new ideas" (p. 26). Essentially, J. P. Prather suggests that in order to advance from unlearning the misconceptions to learning the correct science, students need to test and critique plausible conceptions that they have, as well as those held by scientists. That is to say, students learn the correct science *in part through carefully-guided instruction*.

To engage students in the discovery and articulation processes, one method, as explained by E. E. Prather and Brissenden (2009), is to present a multiple-choice question to the class, and have the students respond with clickers (alternatively "personal response systems" or "personal transmitters"). Optionally students may discuss why they chose their answer with their neighbors in the classroom. Another method is to conduct in-class lecture tutorials, in which students transcribe their ideas and defend them in writing. One drawback to lecture tutorials is that they may be time consuming. Methods that use minimal time in lecture may also be utilized to engage students. For example, here at the University of Maine, Neil F. Comins utilizes two techniques in his introduction to astronomy course. First, Comins concludes each lecture by posing a question about a topic to be addressed in the next class. Students provide handwritten responses at the end of class and turn them in before leaving class. Second, outside of class, Comins allows students to submit their own misconceptions to him through email, as well as misinformation that students identify in other current sources (e.g., in a television show), and explain the correct associated science for small extra credit.

1.2.4 Teaching philosophies

The teaching perspective adopted by J. P. Prather, E. E. Prather and Brissenden (2009), Bailey et al. (2009), and Slater (2013) speaks to the practices typically utilized at one end of a continuum of teaching philosophies. The perspective adopted by these researchers assumes that (i) students already have pre-existing ideas about astronomical phenomena, and (ii) the students bring their ideas to the classroom. In this perspective, termed *constructivism*, students use their pre-existing knowledge to build, refine, or construct more sophisticated knowledge. Bailey and Slater (2004) describe the *constructivist* movement as a “student model-building paradigm” (p. 22), founded on how students synthesize information and experiences into conceptual models. The paradigm is often implemented through some combination of group work interaction and guided tutorials, with minimal involvement on the part of the instructor. In their study, Bailey et al. (2009) refer to “the constructivist movement of the late 20th century” (p. 2), a paradigm in which effective instruction requires building upon pre-instructional student ideas. As emphasized by Bransford (2000), the constructive process may enhance the students’ post instruction understanding of astronomical concepts. Constructivism thus takes place at the intersection of “what students may already know” and “how students can construct more sophisticated models.”

At the opposite end of the continuum of teaching philosophies is *positivism*, which assumes that (i) students are a “blank slate” when they enter the classroom, and (ii) students learn by having the correct science explained to them (Slater, 2013). Traditional, fact-oriented lecture is an example of positivism. As Slater writes, positivism is “the dominant philosophy driving most of teaching.” Unlike constructivism, positivism has no intersection between “what students may already know” and more sophisticated knowledge, because the preconceived knowledge is assumed essentially not to exist. Consequently, the positivist philosophy ignores the likelihood that students may harbor misconceptions before they even enter the classroom. As Bailey et al. note, the outcome of how effectively students learn depends on whether teachers ignore or teach directly to student pre-instructional ideas. If misconceptions interfere with learning, as Smith III et al. (1993) note, and if instructors ignore those misconceptions, then students may be less engaged during instruction, and learning is inhibited.

The method for learning the correct science, as presented by J. P. Prather (1985) and supported by Bailey et al. (2009), Eaton et al. (1983), and E. E. Prather and Brissenden (2009), among other studies, encourages both (i) discovery-based learning and (ii) carefully guided instruction. Discovery-based learning is closer to the constructivist end of the teaching philosophy continuum, while carefully guided instruction is closer to the positivist end of the continuum. In a nutshell, J. P. Prather advocates for *a hybrid of both teaching philosophies*. When discovery-based learning is altogether removed from the classroom, one incorrectly presumes that students learn by being told the information. When discovery-based learning is restored but carefully guided instruction is removed, too much emphasis can be placed on discovery-based learning. As examined by Clark, Kirschner, and Sweller (2012); Moreno (2004, and references therein), one consequence is that “false starts” are common in activities with minimal guidance; students can either become confused and develop misconceptions early in the activity, or the students may claim to have "discovered" a solution without realizing that their interpretation of the information is still incorrect. Clark et al. note that in these minimal guidance settings, "a student is likely to recall his or her *discovery*–not the *correction*" (p. 8).

Here is a brief summary of the various teaching philosophies: The positivist philosophy has the advantage of presenting new information to students in a straightforward and conventional manner, but the disadvantage that preconceived notions that may inhibit learning are assumed essentially not to exist. The constructivist philosophy has the advantage of engaging students in the examination of their own preconceptions and the discovery of alternate conceptions, but the disadvantage that students may develop new misconceptions if they are not strategically guided through the learning process.

1.2.5 An alternate testing instrument

The vast majority of the aforementioned studies rely on recording student responses to a predetermined set of guided questions. For example, in his study, Sadler successfully implemented a 47-item multiple-choice test to examine the nature of misconceptions held by students primarily in high school. Sadler had acquired sufficient knowledge of student misconceptions in astronomy to design questions in the multiple-choice test, with distracter-driven questions that directly target the misconceptions.

While pretests and posttests administered immediately before and after instruction (those which are not part of a longitudinal study) provide meaningful information about short-term retention of information, these tests cannot provide information on the persistence of misconceptions over a longer period. Delayed posttests have been used in several studies within an educational research setting (Lombardi, Sinatrab, & Nussbaum, 2013; E. E. Prather et al., 2004). These tests provide support for conducting studies in educational research in which the data are acquired months after the pretest.

Instead of administering a multiple-choice pretest and posttest, an alternative approach to analyzing student misconceptions in astronomy is to administer a comprehensive inventory of statements, each phrased in the context of a particular misconception, and ask the students to consider each belief directly. One particular retrospective design incorporates the disambiguation of a persistent misconception, either prior to taking the course or at some time during the course. Such a design provides students with the opportunity to give real-time feedback of approximately when, in their lives, they harbored a misconception, or still endorse it even after instruction in the course. The design of such an inventory also permits students to indicate simply if they have never considered one or more of the misconceptions on the inventory. This option is generally not provided on multiple-choice tests, which are also subject to random guessing.

The design of retrospective studies, however, is subject to some issues regarding reliability. Memory is a reconstructive process (Olson & Cal, 1984). As Henry, Moffitt, Caspi, Langley, and Silva (1994) note, a retrospective approach may be of questionable validity in some contexts, notably in the recall of personally-significant emotional and psychosocial material. The authors suggest that “the use of retrospective reports should be limited to testing hypotheses about the relative standing of individuals in a distribution” (p. 92). A comprehensive review by Brewin, Andrews, and Gotlib (1993), however, “suggests that claims concerning the general unreliability of retrospective reports are exaggerated” (p. 82). The likelihood of inaccurate responses may be significantly reduced by asking subjects to provide reports on a timeline for abandoning misconceptions. Hence, in the design of a survey-like instrument, we present brief statements to the students and ask them to respond to the statements directly. Such responses are less likely to be vulnerable to inaccurate self-reports than those associated with the recall of emotionally-significant information across students' lifespans. At the time of this writing, no comprehensive retrospective analysis has been performed on student misconceptions in astronomy. I ,along with dissertation co-advisors Neil F. Comins and Geoffrey L. Thorpe, and David Batuski, have proposed that our newly-designed instrument, which we introduce in Chapter 2, may be a viable option (Favia, Comins, Thorpe, & Batuski, 2014).

1.3 Dissertation organization

This dissertation is organized as follows. In Chapter 2, I discuss the overall structure of the inventory mentioned in Section 1.1, the collection and coding of data, and modifications to the inventory. In Chapters 3, I assess the validity of the inventory as an instrument for examining the retention of misconceptions in astronomy. In Chapter 4, I present an examination of the reliability of student recollections of their own past beliefs, and the extent to which both childhood interest in astronomy and stress influence the recollections. In Chapter 5, I present descriptive statistics on the inventory data. In Chapters 6-12, I interpret the data using principal components analysis, which takes a sample of misconceptions, calculates how strongly each misconception is correlated with the others, and forms inter-correlated sets (or groups) of misconceptions. In Chapters 13-15, I examine these groups using item response theory, which calculates the probability that a member from the population will endorse or reject a misconception, as a function of its tendency to be endorsed. I then show how to use the item response theory results to construct unique orders to teach astronomy topics to both college students and children through adolescents. In Chapter 16, I discuss the design of video lectures and associated pretest/posttest questions to test the orders produced by the item response theory analyses. I conclude with the overall results of this research project in Chapter 17.

CHAPTER 2
THE ASTRONOMY BELIEFS INVENTORY

In this Chapter, I present an outline to the structure of the introductory astronomy lecture course, AST 109, as taught by Prof. Neil F. Comins (NFC) and Prof. David J. Batuski (DJB), here at the University of Maine. I present an overview of a comprehensive list of common incorrect astronomy statements. I discuss how the list was put into the form of an inventory, which was administered to students in AST 109 over six semesters. Then I discuss specific modifications to the inventory for the last two of the six semesters.

2.1 General structure of AST 109

From 2009-2013, the University of Maine introductory astronomy lecture course, AST 109, has been taught by NFC during Fall semesters and by DJB during Spring semesters. Both instructors use *Discovering the Universe* (Comins & Kaufmann III, 2012). The 8th edition was used from Fall 2009 to Spring 2011, and the 9th edition was used from Fall 2011 to Fall 2012. In the Fall 2013 semester, NFC used the 5th edition of *Discovering the Essential Universe* (Comins, 2013); relative to the textbook used previously, *Discovering the Essential Universe* contains updates of information and improved pictures, and most homework questions and ancillary material have been removed. Both instructors cover essentially the same material in the same sequence. Chapters 1-4 of *Discovering the Universe* (9th ed.) present the constellations, the phases of the Moon, gravitation, telescopes, atomic physics, and spectra. Chapters 5-10 present the Nice model, the Earth, the Moon, the properties of each planet from Mercury to Neptune one planet at a time, asteroids, meteoroids, comets, and the Sun. Chapters 11-14 present the life and death of stars and their classifications, as well as the properties of white dwarves, neutron stars, and black holes. Chapters 15-18 present the properties of galaxies and quasars, and finishes with an introduction to cosmology. Chapter 19, while not covered in lecture, gives a brief introduction to astrobiology.

2.1.1 Teaching pedagogy of NFC

In his course, NFC has his students take notes in a published course "skeleton" (Comins, 2008). Topics presented in each lecture are listed in the skeleton in the order presented by NFC, with room between topic headings on each page for the students to take notes. NFC encourages his students to use the skeleton to expedite student note taking, organization of materials, and exam review. In lecture, NFC uses PowerPoint slides to show schematics, pictures, data, and other pertinent from which the students take notes and can ask questions for clarification. During lecture, NFC teaches his students in the context of those particular misconceptions most commonly endorsed by his students, the awareness of which he has developed from his long-term teaching experience (Comins, 2001, 2014). For example, NFC announces common misinformation held by students about a relevant topic to the lecture, and then NFC explains that this information is in fact wrong. An example is provided in the following paragraph. No graded homework is assigned in the course. Extra credit is also awarded to students for identifying their own misconceptions, as well as for identifying misinformation presented in other current sources (e.g., in a television show).

At the end of each class, NFC assigns a misconception-based task, with the topic of the task to be lectured in the subsequent class day preceding the topic. The task is misconception-based, as NFC uses the written responses to inform students of their own misconceptions on the topic. Often the prompt is in the form of a question, and so henceforth I will refer to them collectively as "attendance questions." The responses are handwritten in class and turned in as students leave. Both correct and incorrect answers are awarded equally to the students as attendance credit for that day. For example, if the question is “How many zodiac constellations are there?” then the subsequent lecture would include a discussion about zodiac constellations. (The answer is 13, not 12, as is often commonly thought by his students.) In the subsequent lecture, NFC would present the days of the year for each zodiac constellation, and he would take extra time to point out that (i) the duration of each zodiac is *not* 30 days, and (ii) there are *not* 12 zodiac constellations, with an explanation as to why. NFC uses these attendance questions to aid him in teaching astronomy most directly to the misconceptions. The AST 109 textbook, written by NFC (lead author) and W. J. Kaufman (deceased, 1994), also restates this information. As another example, NFC may ask “What causes the tides on the Earth?” as an attendance question, and in the subsequent class, NFC would explain why the commonly-used picture of the *causes* of one tide directly facing the Moon and the other tide directly opposite the Moon is incorrect. NFC consistently applies the same process to the other attendance questions. Other examples prompt students to describe the causes of the seasons or the rings of Saturn. These prompts were assigned consistently throughout the time frame within which the inventories were administered. An analysis of the data involving roughly a third of the attendance questions asked by NFC in the Fall 2013 semester is presented in Section 4.6.1.

2.1.2 Teaching pedagogy of DJB

In his course, DJB has his students take notes in the same published course “skeleton” (Comins, 2008) used by NFC. Topics presented in each lecture are listed in the order that NFC would present them. DJB employs essentially the same topic sequence as NFC. In his course, DJB assigns traditional homework problems weekly in the form of online quiz questions, to be answered outside of class. Students in the class may earn extra credit by (i) completing a practice final exam, intended to be taken before the actual final exam for preparation, or (ii) attending and reporting on a public talk in astronomy. No extra credit is otherwise awarded to the students.

In lecture, DJB takes attendance by the use of clicker questions. The use of clickers (alternatively "personal response systems" or "personal transmitters") in astronomy has been studied, with results suggesting that the use of clickers engages students in the class, improves their understanding of course concepts, and increases their exam scores (E. E. Prather & Brissenden, 2009). The use of multiple-choice clicker questions whose response options are designed around *a priori* knowledge of common misconceptions held by college students has also been shown to be "an effective method of instruction" (LoPresto & Murrell, 2011, p. 22). Typically, DJB presents these clicker questions sometimes before, but usually after, the topic is covered in the lecture.

The device that students bring to answer clicker questions is a pocket-sized transmitter, with buttons marked for the letters A through E. In lecture, a question with five answers, correspondingly labeled A through E, is presented to the students, who are then given about 30 seconds to select one of the answers. Students receive participation credit in the course by responding (either correctly or incorrectly) to the clicker questions. Some descriptive statistics, such as the percentage of the class who chose each answer, are then presented to the class. Based on the statistics, DJB then decides to move on, if nearly all of the class gets the question correct, or spend more time on the topic if the responses are mixed. Each class contains typically three to five clicker questions, though DJB will occasionally clarify a common misconception in class. The clicker questions "are not usually misconception driven, although . . . misconceptions are frequently involved/probed" (private communication). Table 2.1 presents a sample of clicker questions; these were presented in the Spring 2014 semester. Note that in Table 2.1, Question 4 refers to a model of the relative locations of the stars on a spherical “glass” surface, presented to the students in class for illustration purposes.

1. Which direction has the sun moved on the sky (relative to the stars), since this time yesterday?
2. Which planet revolves opposite to Earth?s revolution?
3. Which planet shows retrograde motion?
4. How many times brighter is a 1.0 magnitude star than a 6.0 mag. ?
5. Why are there no circumpolar constellations at the equator?
6. How many times bigger is our galaxy than our solar system?

Table 2.1. Sample clicker questions

In lecture, DJB uses PowerPoint slides to show schematics, pictures, data, and other pertinent information to the class, during which the students take notes and can ask questions to DJB for clarification. Topics in his lectures are presented from a traditional lecture framework, i.e., one fact at a time. In following this traditional framework, DJB places less emphasis than NFC on explicitly announcing common misinformation held by the students during lecture.

2.2 Development of the ABI

2.2.1 Initial development

Influenced by many studies on education research and misconceptions, as discussed in Chapter 1, Professor NFC developed a preliminary list of 267 incorrect astronomy statements recited by his students. The list, initially developed by NFC, was based on misconceptions identified by NFC between 1985 and 1995 (Comins, 2001). Each item in the list consists of a statement that pertains to a single misconception. NFC administered the list to his students in the form of an inventory in the Fall 1992 and Spring 1993 semesters, with the intent to study when students unlearn (or dispel) astronomy misconceptions. Each topic of the inventory is sequenced to model the topics that were taught by NFC in the course. Because the number of students (18 and 35, respectively for the two semesters) would result in a relatively low sample size, and

NFC had not yet studied the depth of the misconceptions held by his students, I have omitted these data from my research. By 2001, NFC had sufficient data on student misconceptions in all general topics pertaining to astronomy (Comins, 2001) to teach his students in the context of the misconceptions. In 2009, NFC administered the list of beliefs again to his students. Because sample sizes were much larger (~ 100 per semester), enough data were gathered over the span of multiple semesters to develop a research project.

Since Fall 2009, NFC has administered the inventory to students in his class who volunteered themselves, to determine when, in their lives, students in AST 109 disabused themselves of various astronomy misconceptions, or still endorse them. The inventory covers astronomical topics organized in separate sections for the stars, the solar system (which includes a few statements about Jupiter and Pluto), the Moon, the Earth, Venus, Mars, Saturn, the Sun, galaxies, black holes, and misc. general astrophysics concepts. The statements in the inventory are essentially one-liners (e.g., “the Sun is at the center of the Milky Way galaxy,” or “all stars end up as white dwarves”). The statements were organized by topic (eleven in total) and presented in the same sequence as the topics in lecture, except with black holes presented after galaxies in the inventory.

On the last week of classes in each of the Fall 2009 to Fall 2012 semesters, NFC invited students to participate in responding to each statement in the ABI. Students were informed that participation is voluntary, and that by completing the project, extra credit would be added to their course grade as compensation. Extra credit in the course was offered as an incentive for student participation. Student responses in and of themselves had no bearing otherwise on their grades. In the Fall 2013 semester, extra credit was awarded to students who completed both a special pretest for that semester (as discussed in Chapter 4) and the inventory. Students who opted out of any of the assessments were given the opportunity to write an essay based on a topic in astronomy chosen by the instructor for equivalent extra credit. These papers were not used in my research.

The inventory of misconceptions, as administered in Fall 2009, consisted of 267 short statements, nearly all of which are false. NFC reduced the inventory to a set of 235 incorrect statements, starting in the Fall 2010 semester. The revised inventory was administered in the Fall 2010, Fall 2011, and Fall 2012 semesters. In all four semesters, the inventory was titled “AST 109 Misconceptions Workshop.” Students who volunteered were informed that the inventory consists of common statements: they were *not* told that all the statements are wrong. Students were instructed to respond honestly, and that there were no right or wrong answers, so the inventory would have no impact on their grade, other than extra credit for completion. The study was formally approved by the Protection of Human Subjects Review Board at the University of Maine on July 1, 2011, under NFC.

The sets of statements in the inventories, as administered to the students among the four semesters, do not perfectly overlap, as a few misconception-based statements were added to the inventory, while other statements were removed or omitted for various reasons. Altogether, 274 unique statements were archived. Between Fall 2009 and Fall 2010, some statements underwent clarifications in wording that were significant enough to change the context of the statement. The order of the topics common to both the old and the new formats was not changed. Given the desire to have the largest complete sample size, for as many inventory statements as possible, I compromised by electing to base my research on 215 misconceptions that were administrated in common throughout all six semesters. That list of 215 misconception-based statements, organized by topic, is presented in Appendix A. Associated with each statement is a unique label for statement in astronomy (sA). For example, the abbreviation for “statement in astronomy number 106” is “sA106.” There is no significance to the labels other than for identification purposes. Because I archived 274 unique statements, I assigned labels ranging from sA1 to sA274. When the actual inventory was administered to the students, the labels were omitted. The list of 215 items comprises the Astronomy Beliefs Inventory (ABI).

By the Fall 2011 semester, I took on the project of organizing the archived misconception data, from the previous two semesters. On the day of responding to the inventory, students were provided with bubble sheets, on which they recorded their responses to each of the inventory statements. Students were asked to determine if and when they decided it was incorrect. The statements in the ABI are sufficiently universal that only a handful of students over the time frame of the research project had asked, "what if I never believed it?" The directions for indicating their responses, as presented to the students, are provided in Table 2.2. Note that students were also instructed to confirm their response by writing the letter beside each statement on the question sheet.

A)	After the number for each statement please write: **A** if you believed it only as a child **B** if you believed it through high school **C** if you believe it now **D** if you believed it, but learned otherwise in AST 109 If you never thought about a certain statement, please consider it now. Write **E** if the statement sounds plausible or correct to you. Write **F** if you never thought about it before, but think it is wrong now.
B)	If you believe a statement is wrong, please briefly correct it in the space below.

Table 2.2. Directions to responding to the inventory from Fall 2009 through Fall 2012

Students in the Fall 2009 to Fall 2011 semesters were also given the opportunity to correct any incorrect-sounding statements, in writing, on the inventory packet handed to them. Approximately two thirds of the students who volunteered themselves, each semester, made corrections to at least one statement in the inventory.

For the time being, I have not thoroughly investigated the consistency of the majority of the hand-written responses against the responses on the bubble sheets. Corrections to select statements from the inventory can probe the extent to which students consistently interpret the wording of those statements. In Section 3.6, I examine the degree to which students consistently interpret selected statements.

In all semesters except for Fall 2012, the ABI was administered sometime after the last lecture and before the final exam. Due to scheduling constraints and limited student availability for specifically the Fall 2012 semester, administration of the ABI was broken up into two parts. In the Fall 2012 semester, the first part of the inventory was administered prior to the class that covered black holes and cosmology. Students who volunteered themselves were able to respond to 187 statements in the ABI about all other topics covered in the class. The other 28 statements in the ABI were presented to the same students after completion of the final exam. The students who volunteered themselves were reminded to answer honestly to receive the extra credit and were not allowed to refer to their responses on the final exam. Of the students in the Fall 2012 semester who responded to the first 187 items, 96% finished the inventory after taking the final exam.

The response options to the inventory as a whole are designed as an approximate "timeline" for when students abandoned misconceptions in their lives, or if they still retain some misconceptions. For example, a student who responded with **A** or **B** reported having dispelled a misconception sometime during childhood or adolescence, which is before the student took AST 109. A student who responded with **D** or **F** was either never aware of the misconception, but decided that the statement was incorrect, or learned that the statement was in fact false as a result of taking AST 109. A student who responded with **C** or **E** endorsed the misconception, even after instruction on the topic in the course. A discussion on how the responses are coded is presented in Section 2.2.3.

The data from the first format of the ABI, as described in Table 2.2, was not analyzed until after the ABI was administered in the Fall 2011 semester. The format was limited in that (i) all of the statements were presented as false, and (ii) statements were grouped together by topic. It is perhaps an unfortunate oversight that we had not considered the extent to which variations in the format of the ABI may have affected the data earlier in the time frame of the overall research project. In Section 2.2.2, I present modifications to the ABI for administration beginning in the Spring 2013 semester.

There is, however, substantial promise that the data in the Fall 2009 to Fall 2012, in which the first format of the ABI was used, can provide meaningful insights regarding student misconceptions. For example, in Chapter 7, I argue that the format of the ABI has no significant effect on *correlations* between misconceptions within topics of the ABI. In Chapter 9, I use available data from the two different ABI formats to present explicitly the nature of correlations regarding misconceptions in a couple of the topics of the ABI. I use these analyses to show that the data from the previous ABI format can in fact be salvaged despite the all-false nature of the statements.

2.2.2 Inventory modification

A special format of the inventory was used for the Spring 2013 and Fall 2013 semesters. In this alternate format, the changes indicated in Table 2.3 were made to the inventory. The modified format was used to test the reliability of the inventory that was administered in the previous four semesters. In the Spring 2013 and Fall 2013 semesters, students were not asked to correct any incorrect statements.

The new response option format, presented in Table 2.3, gives students the opportunity to decide, *first*, if each statement is true or false, and *then* reflect on when they learned it. For the purposes of my analysis, a false statement is a statement phrased as an incorrect belief that pertains to a specific misconception (e.g., sA111,

1. 86 of the 215 statements from the inventory were turned from false to true, and the statements were called *beliefs*.
2. The sequence of the statement presentation was randomized.
3. The response option format was modified to the format below.
4. Students were not asked to correct any incorrect statements.

For each statement, ***first decide if the statement is true or false****.*
After you have decided:

If you think the statement is ***true****, enter:*

A if you learned this before high school,
B if you learned this in high school,
C if you learned this in AST 109,
D if you never considered this statement before today.

If you think the statement is ***false****, enter:*

E if you learned this before high school,
F if you learned this in high school,
G if you learned this in AST 109,
H if you never considered this statement before today.

Table 2.3. Changes to the ABI format for the Spring 2013 and Fall 2013 semesters

"Earth's axis is not tilted compared to the ecliptic"), and a true statement is a statement that is scientifically accurate (e.g., "Earth's axis is tilted compared to the ecliptic). The scientifically-accurate statements are entirely based off previously-incorrect statements, just with the opposite meaning. In contrast, many incorrect beliefs (e.g., sA70, "comets are molten rock hurtling through space at high speeds and their tails are jet wash behind them") cannot be concisely phrased with the opposite connotation, other than to insert "not" at the beginning. I have elected to change no more than 86 statements from false to true so as to minimize the insertion of "not" repetitively throughout the modified inventory.

In addition to the re-phrasing of statements, the order of the statement presentation was also randomized into three different unique sequences of the same statements, so that separate inventory forms were created, one for each unique sequence. All forms still presented the same beliefs as those from the Fall 2009 to Fall

2012 semesters. One of the random orders, presented in Appendix B, includes the 215 statements under consideration in this study. Futhermore, in the Spring 2013 and Fall 2013 semesters, some statements were turned from false to true. These are marked with "n" (for "negated") after the identifier. For example, "sA182: all galaxies are the same in size and shape" was changed to "sA182n: galaxies have a variety of sizes and shapes." Again, the labels starting with "sA" were omitted from the actual inventories presented to the students.

2.2.3 Coding the data

A master coding scheme was developed for all responses to the first and second formats of the inventories. The motivation for the codes (defined momentarily) is the desire to preserve the sense of "timeline," or a *rank* as to approximately how late in one's life does one abandon a misconception. Namely, the timeline inferred from the codes is "before AST 109," "during AST 109," and "after AST 109." Motivated by this timeline, I developed a coding scheme consisting of three relative degrees of misconception *retainment*, where I define retainment as the tendency for students to hold on to a misconception from either their childhood or during some point in the course. A code of "1" means a student disabused oneself of a misconception as a child or adolescent and so indicates the lowest relative degree of misconception retainment. A code of "2" means a student may have harbored a misconception but unlearned or otherwise disabused oneself of it by the end of the course. A code of "3" means a student still believes the misconception, which indicates the highest relative degree of misconception retainment. These codes are summarized in Table 2.4. In Chapter 5, I present descriptive statistics of the ABI responses as a whole, and I present the results of several tests of the reliability of the ABI responses.

In Table 2.4, a code of "1" means a student disabused oneself of a misconception as a child or adolescent. That students may report and then recall

1:	unlearned the incorrect belief as a child or adolescent, indicating the lowest degree of misconception retainment
2:	unlearned the incorrect belief as a result of taking AST 109, indicating a medium degree of misconception retainment
3:	retained the incorrect belief even after instruction in AST 109, indicating the highest degree of misconception retainment

Table 2.4. Codes for three relative degrees of misconception retainment

disambiguation of a misconception as far back as one's own childhood may prompt a criticism as to whether or not students are providing accurate reports of their own beliefs. As I show in Chapter 3, however, there is little concern for the accuracy of these reports, because the accuracy of reports on the ABI is comparable to that of instruments designed in a more traditional multiple-choice format. Also, the data that I collect are subject to a correlation analysis, as I discuss in Chapters 6-4. The purpose of this analysis is to determine correlations between misconceptions. In other words, if students endorse a particular misconception on the ABI, then I can determine how likely students are to endorse related misconceptions.

Note that in the Spring 2013 and Fall 2013 semesters, 86 of the 215 statements under consideration were changed from incorrect to scientifically accurate, as discussed in Section 2.2.2. Hence, for a scientifically accurate statement such as "the Milky Way is one of many galaxies" (associated with sA218), if a student believed this while a child or adolescent, then the student's response was coded "1." If the student learned this from taking the course, then the student's response was coded "2." If the student did not believe the correct statement, then the student's response was coded "3." This procedure applied to the remaining scientifically-accurate statements in the revised format of the ABI.

2.3 Sample sizes

Students were asked to respond to each statement on bubble sheets with the corresponding letters for each response option. Bubble sheets were also checked by the instructor upon student completion of the survey, to ensure that the proper number of bubbles were filled in. The bubble sheet responses prevailed over hand-written responses, as were available from archived materials for the Fall 2009 to Fall 2011 semesters, when no issue was observed on the bubble sheets. Bubble sheets were checked by eye whenever a response was scanned and marked invalid, e.g., if a student changed an answer. Two students in the Fall 2012 semester were removed from the final sample, because they did not complete the inventory.

The total participant count for all six semesters $N = 639$, of which 341 students identified themselves as male, and 297 students identified themselves as female. Demographic information is available for all but one student. The overall age of the sample is 20.0 ± 3.8 years, though the distribution is heavily skewed. The minimum age is 17, and the maximum age is 62. Seven students were at least of age 40, and 30 students were at least of age 25. Respectively, the percents of subjects whose ethnicities are caucasian, native american, hispanic, asian, black, and other/unspecified are 84.6%, 2.0%, 1.9%, 1.4%, 0.9%, and 8.8%. Table 2.5 presents the sample size for each semester of AST 109 in which the ABI was administered. On average, about 60% of the class took the ABI.

Semester	**Instructor**	**Class Size**	**Sample Size**
Fall 2009	NFC	188	114
Fall 2010	NFC	175	107
Fall 2011	NFC	171	91
Fall 2012	NFC	170	91
Spring 2013	DJB	192	126
Fall 2013	NFC	174	110

Table 2.5. ABI sample and class sizes for each semester

CHAPTER 3

ADDRESSING CRITICISMS OF THE ASTRONOMY BELIEFS INVENTORY

The ABI is an instrument originally designed by NFC to assess when, in the lives of students, they unlearned various misconceptions in astronomy. Responses to the ABI partly depend on accurate self-reports. The ABI is also a rather lengthy instrument, in which students are asked to provide accurate self-reports of 215 statements. In this Chapter, I address potential related criticisms about the ABI.

3.1 Mean misconception retainment score

The three relative degrees of misconception retainment, in Table 2.4, provide a scale with which I can estimate the relative tendency for students to endorse the misconception associated with a statement in the ABI. A misconception retainment score of 1 indicates the lowest degree of misconception retainment, in that the associated misconception is dispelled prior to college. A misconception retainment score of 3 indicates the highest degree of misconception retainment, in that the associated misconception is endorsed even after instruction. Hence, the misconception retainment score for any particular statement can be no lower than 1 and no higher than 3.

The coding scheme that I have introduced serves as a rank. Similar coding schemes are used to make comparisons, most notably, on a Likert-type scale (Miller, Lovler, & McIntire, 1999, p. 117). While taking a survey whose items use a Likert-type scale, a subject indicates a level of agreement or disagreement with a reflective statement (e.g., “I like to go out when it is sunny”). Typical options for these statements are: “strongly disagree,” “disagree,” “neither agree nor disagree,” “agree,” and “strongly agree.” That the options are ranked supports treating the data as one would on an ordinal scale; however, as Miller et al. notes, professionals sometimes treat the data on an equal interval scale to make comparisons between groups. On such a scale, the difference between a "1" and a "2" carries the same context as the difference between a "2" and a "3." While the codes that I established in Table 2.4 do not rigidly meet this criterion, neither do Likert-type scales, since the metric for each successive level of agreement (or disagreement) is based on the available response options to one's personal reflection of the statement. Nonetheless, coding schemes can be developed for either scale to make some quantitative comparisons, such as frequency, mean, and standard deviation, and so I will proceed analogously using my coding scheme for relative measures of misconception retainment.

The mean misconception retainment score (MMRS) for each student is the mean over the responses to all statements: A mean misconception retainment score of much greater than 2 indicates that the student retains misconceptions in astronomy as a whole, even after instruction. A mean misconception retainment score of near 2 indicates that the student tends to hold on to misconceptions in astronomy as a whole until instruction in AST 109. A mean misconception retainment score of much less than 2 indicates that the student tends to come into AST 109 having already disabused oneself of misconceptions in astronomy as a whole. To calculate the mean misconception retainment score for each student, I sum over the misconception retainment scores (1, 2, 3) for each item on the inventory, then I divide the result by the number of items to which the student responded. For example, if a student responded to only 212 of the 215 statements, then the scores for the 212 statements were summed, and the total was divided by 212 to obtain the mean misconception retainment. Of the total sample, 89% of the students responded to all 215 statements, and 98% of the students responded to at least 212 statements.

The MMRS provides a relative measure of the overall tendency for each student to retain misconceptions in astronomy. However, some statements in the ABI may be associated with misconceptions that are harder to dispel, even after instruction.

To assess the relative difficulty of the statements, for each statement, I calculated a separate mean using all of the student responses, coded as degrees of misconception retainment, to that statement. Statements with a higher overall degree of misconception retainment are associated with misconceptions that are harder to dispel. The overall degree of misconception retainment, for each of the 215 statements, is reported in Appendix A. For example, the overall degree of misconception retainment for sA1, "all of the stars were created at the same time," is 1.60, whereas the overall degree of misconception retainment for sA2, "there are 12 zodiac constellations," is 2.13, indicating that the misconception associated with sA2 is, on average, harder for students to dispel than the misconception associated with sA1.

3.2 Brief review of univariate statistics

Because the ABI was administered to students in six semesters, one may expect that the MMRS may vary from one semester to another. The extent of the differences among the semesters is quantified by using an ANalysis Of VAriance (ANOVA) test, described here. In an ANOVA test, one compares the mean of the scores in a group (a unique sample of the available data) with the mean of the scores within other groups. The way that means are compared in an ANOVA test is by comparing the variances (the variance being the square of the standard deviation) in the scores. The ratio of the variance of the scores between groups vs. the variance of the scores within groups is called the F-statistic or F-ratio. Within the F-ratio are two quantities for the degrees of freedom, using the notation $F(df_1, df-2)$. The first argument, df_1, also called the *between-group* degrees of freedom, is one less than the number of groups being compared. The second argument, df_2, also called the *within-group* degrees of freedom, is one less than the difference between the total sample size and df_1. For example, $F(5, 633)$ indicates that there are $5 + 1 = 6$ groups whose means are being compared, and the total sample size is $5 + 1 + 633 = 639$.

A key presupposition of the ANOVA test is that the groups share equal variances. The validity of this assumption is quantified by performing Levene's test of homogeneity of variances. Among the various ways of quantifying the homogeneity of variances among groups, the Levene's test is a relatively robust test of the deviation from homogeneity of variances in the scores among the groups (Borkowski, 2014; Hole, 2013). Levene's test produces (i) a Levene statistic (which I will designate W), which measures the deviation in the homogeneity of variances among the groups, and (ii) the significance in the differences in the variances (which I will designate p_V, so as to not be confused with the usual p statistic for comparing actual group means). The value of the p_V statistic determines whether or not the variance in one group is significantly larger or smaller than the variance in any of the other groups. The p_V statistic does not compare the means of the groups

Provided that one can demonstrate the presupposition that the variances among the scores do not differ significantly, the meaning of the ANOVA test results can be understood as follows: a statistically significant result indicates that the mean of the scores within at least one group differs significantly from the mean of the scores within the other groups. A result is "significant" if, in an ANOVA test, the mean of a group falls outside of the range of values estimated on the assumption that the data are normally distributed. In statistics, one often assumes a 95% confidence interval, which is to say that a result is "significant" if it falls outside of where 95% of the data would be expected under the assumption of normality. In other words, the result is "significant" if the probability of obtaining that result, or something more extreme, by chance, is no more than 5%, or $p < 0.05$. These values are significant, because they indicate that there is a one in 20 chance of drawing the conclusion that the means are significantly different, and being incorrect. (The choice of the size of the confidence interval, e.g., 95%, is solely a convention.) For example, if the MMRS for a particular semester is 1.70, and the MMRS for another semester is 1.84, then one cannot determine if these means are significantly different from each other until one performs a test of significance (e.g., an ANOVA test). Further discussion of tests of significance are presented in additional resources (Lacey, 1998; Walker, 1985).

An additional statistic is necessary to assess the tendency to obtain a significant result given the limitation of the available sample size (Walker, 1985, pp. 348-349). The magnitude of the difference among the means for each group, relative to the overall variation in the data, is called the *effect size*. A small effect size represents a significant effect that is more likely to be detected with a larger sample size, whereas a large effect size represents a significant effect that is relatively easy to detect even with a smaller sample size. In an ANOVA test, several measures are used to calculate the effect size (Becker, 1999; J. Cohen, 1988), of which η^2 and ω^2 are summarized here. A common estimation of the effect size is given by

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}} \tag{3.1}$$

where $SS_{Between}$ is the sum of squares (that is, for a data point X_i and a mean $\bar{X}$ in a group, $SS = \sum(X_i - \bar{X})^2$) between individual groups, SS_{Within} is the sum of squares within the groups, and $SS_{Total} = SS_{Between} + SS_{Within}$. The estimate η^2 of the effect size is considered “small” for $0.010 < \eta^2 \leq 0.059$, “medium” for $0.059 < \eta^2 \leq 0.138$, and “large” for $\eta^2 > 0.138$. Because η^2 measures only the sample and not the actual population, one may use a less-biased effect size estimator, ω^2, given by

$$\omega^2 = \frac{SS_{Between} - df_{Between} \times MS_{Within}}{SS_{Total} + MS_{Within}}, \tag{3.2}$$

where

$$MS_{Within} = \left(\frac{SS_{Within}}{df_{Within}}\right) \tag{3.3}$$

is the mean square of the data within groups, and df is the degrees of freedom. Since ω^2 accounts for variation within groups, rather than just between groups, ω^2 is less biased than η^2. It is possible, however, that for some non-significant results, the numerator in the definition of ω^2 may become negative, which essentially means that the effect size is negligibly small. Throughout my dissertation, I will refer to η^2 and ω^2 to estimate effect sizes.

3.3 Effect of statement presentation order

In Section 2.2, I outlined two unique formats of the ABI. The first format consisted of topics that were presented essentially in the same order as the material in the course. The modified format was presented to the students in the Spring 2013 and Fall 2013 semesters, with the order of the statements randomized to address criticisms related to the proximity of statements and the reliability of self-reports toward the beginning and end of the study.

In the Spring 2013 semester, of the three different orders of statements, 42 students received form #1, 43 students received form #2, and 41 students received form #3. In the Fall 2013 semester, 36 students received form #1, 36 students received form #2, and 38 students received form #3. The data were coded in the manner presented in Table 2.4. I performed an ANOVA test on the three different forms (e.g., form #1 represents one group). Table 3.1 and Table 3.2 present the significance of differences in overall degree of misconception retainment for each of the three forms, in the Spring 2013 and Fall 2013 semesters. Included in Table 3.1 and Table 3.2 are the mean (and associated standard deviation) of misconception retainment for each topic, the Levene statistic W, the significance of the difference in variances p_V among groups, the F-ratio, and the significance p of the difference in means among groups. Values of $p < 0.05$ are considered statistically significant, in that the differences in scores among the forms are unlikely to come by chance, and these values are marked with an asterisk

(*). Note that the section labeled Three Planets combines all statements pertaining to Venus, Mars, and Saturn, in their respective sections of the inventory (refer to Appendix A). For the Spring 2013 and Fall 2013 semesters, respectively, $n = 126$, and $n = 110$. The range of all possible scores is from 1 to 3 for all tests, as these are the codes that represent the relative degrees of misconception retainment.

Under "Spring 2013," in Table 3.1 and Table 3.2, I have indicated the significance of the Moon topic with an asterisk, because $p = 0.022$ is a statistically significant result. Meanwhile, statements normally under the Stars, Solar System, and other topics of the ABI did not exhibit any significant variation in misconception retainment scores, so one significant result out of the nine topics is not out of the ordinary. For all 215 statements, the variation in misconception retainment scores for the Spring 2013 semester did not differ significantly among the three forms ($p = 0.081$), indicating that *presentation order had no significant effect on student misconception retainment scores for the ABI*. Likewise, the variation in misconception retainment scores for the Fall 2013 semester did not at all differ significantly among the three forms ($p = 0.903$), indicating that *presentation order had even less of an effect on student misconception retainment scores for the ABI*.

With regard to the data for both the Spring 2013 and Fall 2013 semesters, Table 3.1 and Table 3.2 show that there was generally no statistically significant difference in the ABI scores among the three forms, indicating that the proximity of neighboring statements had no significant influence on student responses.

3.4 Effect of fatigue

As administered to the students in the Fall 2010, Fall 2011, Fall 2012, Spring 2013, and Fall 2013 semesters, the full survey contained 235 items. Students typically spend between one and one and a half hours responding to all 235 items. A question that may be raised is whether or not student fatigue, at some point during the response

	Spring 2013							
ABI Topic	**Mean**	**Std. Dev.**	W	p_V	$F(2,123)$	p	η^2	ω^2
Stars	1.93	0.25	0.474	0.624	1.403	0.250	0.022	0.006
Solar System	1.86	0.32	1.261	0.287	3.048	0.051	0.047	0.032
Moon	1.79	0.28	0.367	0.693	3.935	0.022*	0.060	0.044
Three Planets	1.81	0.32	0.669	0.514	1.450	0.239	0.023	0.007
Earth	1.92	0.27	0.782	0.460	1.320	0.271	0.021	0.005
Sun	2.00	0.27	0.657	0.520	3.026	0.052	0.047	0.031
Galaxies	1.83	0.32	1.994	0.141	1.929	0.150	0.030	0.015
Black Holes	2.01	0.29	0.944	0.392	1.990	0.141	0.031	0.015
General Astrophysics	2.10	0.25	0.046	0.955	1.528	0.221	0.024	0.008
All 215 Statements	**1.91**	**0.25**	**0.625**	**0.537**	**2.567**	**0.081**	**0.040**	**0.024**

Table 3.1. Significance of differences among the three random sequences of ABI statements for the Spring 2013 semester, with terms as defined in the text

			Fall 2013					
ABI Topic	**Mean**	**Std. Dev.**	W	p_V	$F(2,107)$	p	η^2	ω^2
Stars	1.93	0.26	0.107	0.898	0.036	0.965	0.001	–0.018
Solar System	1.87	0.30	1.744	0.180	0.450	0.639	0.008	–0.010
Moon	1.81	0.29	0.353	0.704	0.279	0.757	0.005	–0.013
Three Planets	1.77	0.33	0.522	0.595	0.706	0.496	0.013	–0.005
Earth	1.88	0.26	0.306	0.737	0.065	0.937	0.001	–0.017
Sun	2.00	0.25	1.199	0.305	0.244	0.784	0.005	0.003
Galaxies	1.82	0.35	0.313	0.732	0.001	0.999	0.000	–0.019
Black Holes	2.10	0.30	0.065	0.937	0.094	0.910	0.002	–0.017
General Astrophysics	2.07	0.26	0.816	0.445	0.526	0.593	0.001	–0.009
All 215 Statements	**1.91**	**0.25**	**0.126**	**0.882**	**0.102**	**0.903**	**0.002**	**–0.017**

Table 3.2. Significance of differences among the three random sequences of ABI statements for the Fall 2013 semester, with terms as defined in the text

process, interferes with the validity of responses provided by the students thereafter. Fatigue could well be a factor when completing a lengthy inventory. If so, then one would expect that responses to the ABI become less meaningful in the later sections. If this is true, then the correlations between earlier and later items would be low. Given that the order of statement presentations was randomized in the Spring 2013 and Fall 2013 semester, one can use data from these semesters to see if there is any difference between early and later responding, regardless of item content.

To examine the extent to which students become fatigued in the process of completing the ABI, I used data from all of the original 235 statements in the inventories administered to the students in the Spring 2013 and Fall 2013 semesters. In these semesters, the sequence of the 235 statements was randomized (refer to Section 2.2). The data were coded in the manner presented in Table 2.4. As a first measure, I calculated two mean misconception retainment scores, one for each of the first and second halves of the inventories (respectfully, 118 and 117 statements) as presented to the students. Since the three formats contain statements in essentially random orders, the scores on the first and second halves are expected to be well correlated. The correlation of responses between each pair of items a and b is calculated by the sample correlation coefficient C_{ab}, given by

$$C_{ab} = \frac{\sum_{i=1}^{N}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{N}(a_i - \bar{a})^2 \times \sum_{i=1}^{N}(b_i - \bar{b})^2}}, \tag{3.4}$$

where a_i and b_i respectively represent the scores on items a and b for student i, and $\bar{a}$ and $\bar{b}$ are the mean scores of items a and b, respectively. For the Spring 2013 semester, where $n = 126$, I report summary statistics for the first half ($M = 1.88$, $SD = 0.26$) and the second half ($M = 1.95$, $SD = 0.26$), where M and SD are the mean and standard deviation, respectively. For the Fall 2013 semester, where $n = 110$, I analogously report summary statistics for the first half (M = 1.87, SD = 0.25) and the second half (M = 1.96, SD = 0.26). Table 3.3 summarizes the correlations between the first and second halves. Values of $p < 0.0005$ are considered statistically significant, in that the differences in scores among the forms are unlikely to come by chance, and these values are marked in Table 3.3 with a double-asterisk (**). The meaning of r^2, the square of the correlation coefficient, will be discussed momentarily.

Semester	***n***	**Correlation**	**r^2**	***p***
Spring 2013	126	0.772	0.60	<0.0005**
Fall 2013	110	0.864	0.75	<0.0005**

Table 3.3. Correlation between scores on the first and second half of the ABI

The data in Table 3.3 show that for the Spring 2013 semester, misconception retainment scores in the first half of the inventory were correlated (coefficient of 0.772, $p < 0.0005$) with misconception retainment scores in the second half of the inventory. That the correlation is very significant suggests that the way that students responded to the statements remained consistent from start to finish. Likewise, the data in Table 3.3 show a similar correlation for the Fall 2013 semester (coefficient of 0.864, $p < 0.0005$), which further supports that student responses remain consistent from start to finish. On the basis of this analysis, there is no evidence that students respond differently between the first and second half of the inventory, which is consistent with the hypothesis that fatigue does not sacrifice the validity of the data.

When interpreting the statistics of a correlation coefficient (e.g., 0.772 for the Spring 2013 semester), one can report the significance (the p-value), which indicates whether or not the correlation is different from zero. The significance, however, depends on the sample size. An additional interpretation, given by R. Taylor (1990), uses the *square* of the correlation, also known as the *coefficient of determination* (or r^2 value, where r is the correlation coefficient). According to R. Taylor, the coefficient of determination gives a more meaningful interpretation of correlations between variables, because r^2 reports the total variation in one variable that can be explained (or accounted for) by the variance in the other. According to Table 3.3, for the Spring 2013 semester, $r = 0.772$, so $r^2 = 0.60$, which says that 60% of the variance in the second half is accounted for by the variance in the first half.

As an additional check on the influence of fatigue (if any) on student responses to the inventory, I analyzed the internal consistency of the responses to select topics within the ABI. The internal consistency of a set of data is reported by coefficient alpha (α), sometimes called Cronbach's alpha, which is a measure between 0 and 1 indicating how closely related a set of variables are in a group (Schmitt, 1996, and references therein). The value $\alpha \approx 0$ indicate highly dissimilar items, whereas the value $\alpha \approx 1$ indicates extremely similar items, with $\alpha = 1$ referring to two or more identical data sets. Values of $\alpha \geq 0.70$ preliminarily represent a group of items with "adequate" internal consistency; although as Schmitt notes, coefficient α is a function of the length of a test (with K items). Coefficient α is given by

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_i^2}{\sigma_K^2}\right), \tag{3.5}$$

where σ_K^2 is the variance of all responses for all K variables. In an example presented by Schmitt, for $K = 10$, if the correlation between any one item and any other item in the set is 0.3, then $\alpha = 0.81$. If a longer test yields the same pattern and value of item inter-correlations, then α would increase.

Using the original 235 statements, as administered to the students, I calculated α of the misconception retainment scores from the Spring 2013 and Fall 2013 semesters, separately for the Earth topic, with 37 statements, and the Sun topic, with 32 statements. Since the three formats contain statements in essentially random orders, the internal consistency in scores among the random orders should be essentially the same. For the Spring 2013 semester, I report summary statistics for the Earth topic

(M = 1.92, SD = 0.27) and the Sun topic (M = 2.00, SD = 0.27); these are the same as in Table 3.1. For the Fall 2013 semester, I analogously report summary statistics for the Earth topic (M = 1.88, SD = 0.26) and the Sun topic (M = 2.00, SD = 0.26). Table 3.4 presents α for the three orders, as administered to students in the Spring 2013 semester and the Fall 2013 semester.

Spring 2013

Form	n	α **(Earth)**	α **(Sun)**
1	42	0.79	0.74
2	43	0.84	0.80
3	41	0.83	0.81

Fall 2013

Form	n	α **(Earth)**	α **(Sun)**
1	36	0.83	0.80
2	36	0.84	0.85
3	38	0.84	0.78

Table 3.4. Internal consistency of student responses to the Earth and Sun topics in the Spring 2013 and Fall 2013 semesters

A brief interpretation of the values of coefficient α in Table 3.4 follows. While α may seem low given the large number of items per topic, not all items within a topic strongly inter-correlate with each other. Statements within each topic of the ABI are associated with a particular factor structure that describes the inter-item correlations. In Chapter 10, I subdivide the statements within each topic into various groups determined using factor analysis, which establishes the groups based on highest inter-item correlations. The statements within each group exhibit relatively high correlations; however correlations between statements of different groups tend to be relatively low. Hence, not every item inter-correlates strongly with every other item in the topic. The same is true for all other topics in the ABI. What is of importance here is that the response data have at least adequate internal consistency. It is thus unlikely that student fatigue would threaten the internal consistency of the inventory data.

3.5 Effect of false vs. true statements

As discussed in Section 2.2.2, I introduced a modified inventory format, in which 86 of the original 215 statements were changed. Originally, the statements were false, in the sense that each statement was phrased as an incorrect belief which pertains to a misconception. The 86 true statements were taken from the pool of incorrect beliefs and changed into scientifically-accurate statements. Essentially, 86 statements were changed from “false” to “true,” and I will use this interpretation of false and true henceforth when I refer to particular statements.

3.5.1 Analysis of all 215 statements

In the Spring 2013 and Fall 2013 semesters, I used the modified ABI format, as illustrated in Table 2.3. In the modified format, I changed 86 of the 215 statements from false to true. In the context of those statements which I changed from false to true, an “incorrect belief” is one in which students reject a truthful statement. For example, if a student rejects sA255n, “light travels at a finite speed,” then the student harbors *a misconception associated with a true statement*. This is in contrast to if a student endorses sA255, “light travels infinitely fast,” in which the student harbors *a misconception associated with a false statement*.

Because the first format of the ABI consists of essentially falsely-worded statements, one criticism that may be raised is that students who take the inventory may assert early in the response process, from the context of the statements, that the statements are all false, which may introduce a bias in student responses thereafter. To examine the extent of such a bias, I calculated the mean fraction, over all statements in the ABI, of incorrect statements that students believed even after taking AST 109, for each of the six semesters. To calculate the fraction of misconceptions endorsed even after instruction in the course, I took all ABI data, coded as degrees of misconception retainment as described in Table 2.4, and recoded the data into two categories: one for endorsing the incorrect belief even after instruction, and one for unlearning the incorrect belief anytime before the end of the course. Table 3.5 presents a summary of the recoded responses, where a score of "0" indicates that a misconception was dispelled during or prior to instruction, and a score of "1" indicates that a misconception was endorsed even after instruction. *The proportion of 1s represents the fraction of misconceptions endorsed.*

Degree of Misconception Endorsement	Endorsed Misconception
1 →	→ 0
2 →	→ 0
3 →	→ 1

Table 3.5. Recoded responses indicating endorsement of a misconception after instruction in AST 109

The values in Table 3.6 are drawn from all 215 statements, scored to indicate the endorsement of incorrect beliefs. The mean and standard deviation for each semester is reported; however, the standard deviation should not be confused as a measure for comparing differences in the means between semesters. Such differences are appropriately compared by the use of tests of the significances of the means, such as an ANOVA test for comparing the means among multiple data sets, as discussed in Section 3.2, or a paired samples t-test for comparing the means between two sets, as described later in Section 3.5.4. These tests provide better indications of the significances of differences *between* means than do standard deviations alone, because standard deviations are useful for measuring the spread of data *within a particular set*.

Semester	***n***	**Mean Fraction Believed**	**Std. Dev.**
Fall 2009	114	0.201	0.109
Fall 2010	107	0.173	0.109
Fall 2011	91	0.186	0.109
Fall 2012	91	0.117	0.097
Spring 2013	126	0.275	0.105
Fall 2013	110	0.252	0.101

Table 3.6. Mean fraction of incorrect beliefs endorsed per semester, using all 215 statements

The *variances* in the mean fractions of misconceptions endorsed among all six semesters (including Spring 2013 and Fall 2013) were not significantly different from each other ($W = 1.093$, $p_V = 0.363$). This is to say, the *variances* (or the squares in the standard deviations of the two semester sets) do not differ significantly from each other. An ANOVA test confirms that among the semesters, the *mean fractions* of incorrect beliefs endorsed even after instruction differ significantly from each other ($F(5{,}633) = 30.75$, $p < 0.0005$, $\eta^2 = 0.195$, $\omega^2 = 0.189$). Table 3.6 presents the mean fraction of incorrect beliefs endorsed even after instruction per semester, and the standard deviation of the mean fraction of incorrect beliefs endorsed.

Inspection of the data in Table 3.6 shows that the values for the Spring 2013 and Fall 2013 semesters are numerically higher than those of the first four semesters (Fall 2009 to Fall 2012), which may account for the observed significance in misconception endorsement scores among the six semesters. In the first four semesters, the ABI was administered in the format described in Table 2.2. In the Spring 2013 and Fall 2013 semesters, the ABI was administered in the revised format, as described in Table 2.3, in which 86 of the 215 statements were phrased as true.

To test whether or not the scores between the two semester sets are significantly different from each other, I performed an ANOVA test on the mean fraction of incorrect beliefs endorsed between the two sets. The variances in the MMRS between the two semester sets were not significantly different from each other ($W = 1.953$, $p_V = 0.163$). An ANOVA test confirms that between the semester groups, the mean fractions of incorrect beliefs endorsed even after instruction differ significantly from each other ($F(1{,}637) = 110.4$, $p < 0.0005$, $\eta^2 = 0.148$, $\omega^2 = 0.146$). The result of this ANOVA test suggests that the format of the ABI may play a significant role on the overall reported degree of misconception endorsement even after instruction in the course.

3.5.2 Analysis of 129 always-false statements

As a further examination of student responses between the different ABI formats, I created two reduced data sets: one consisting of exclusively those 129 incorrect statements, that is, those whose wording throughout all six semesters remained false, and one consisting of exclusively those 86 statements which I reworded to be scientifically accurate, for just the Spring 2013 and Fall 2013 semesters. With regard to the reduced data set of 129 false statements, the variances in the mean fractions of incorrect statements believed even after instruction among the six semesters were not significantly different from each other ($W = 1.604$, $p_V = 0.157$). An ANOVA test confirms, once again, that among the semesters, the mean fractions of incorrect statements believed even after instruction differ significantly from each other ($F(5,633) = 23.59$, $p < 0.0005$, $\eta^2 = 0.157$, $\omega^2 = 0.150$). Table 3.7 presents the mean fraction of the 129 incorrect statements believed even after instruction, per semester, and their standard deviations.

Semester	***n***	**Mean Fraction Believed**	**Std. Dev.**
Fall 2009	114	0.224	0.115
Fall 2010	107	0.197	0.119
Fall 2011	91	0.206	0.120
Fall 2012	91	0.133	0.102
Spring 2013	126	0.302	0.136
Fall 2013	110	0.264	0.140

Table 3.7. Mean fraction of incorrect beliefs endorsed per semester, using exclusively 129 false statements of the original set of 215 statements

The values in Table 3.7 are drawn from 129 false statements, whose wording remained false even after changing the inventory format (as described in Table 2.3) for the Spring 2013 and Fall 2013 semesters. Inspection of the data in Table 3.7 shows that the values for the Spring 2013 and Fall 2013 semesters are again numerically higher those of the first four semesters (Fall 2009 to Fall 2012), which may account for the observed significance in misconception endorsement scores among the six semesters.

To test whether or not the scores between the two semester sets are significantly different from each other, I performed an ANOVA test on the mean fraction of incorrect beliefs endorsed between the two sets. The variances in the mean fraction of incorrect beliefs endorsed between the sets were not significantly different from each other ($W = 2.87$, $p_V = 0.091$). An ANOVA test confirms that between the semester groups, the mean fractions of incorrect beliefs endorsed even after instruction differ significantly from each other ($F(1,637) = 78.85$, $p < 0.0005$, $\eta^2 = 0.110$, $\omega^2 = 0.109$), which leads one to make the same suggestion as in Section 3.5.1, that the format of the ABI may play a significant role on the overall reported degree of misconception endorsement even after instruction in the course.

3.5.3 Analysis of 86 statements re-phrased as true in the Spring 2013 and Fall 2013 semesters

With regard to the reduced data set of 86 statements which were reworded to true for the Spring 2013 and Fall 2013 semesters, the variances in the mean fractions of correct statements rejected even after instruction among the six semesters were not significantly different from each other ($W = 1.551$, $p_V = 0.172$). An ANOVA test confirms, once again, that among the semesters, the mean fractions of incorrect statements believed even after instruction differ significantly from each other ($F(5,633) = 33.41$, $p < 0.0005$, $\eta^2 = 0.209$, $\omega^2 = 0.202$). Table 3.8 presents (i) the mean fraction of the time that students believed the incorrect idea associated with those 86 statements which were phrased as true for the Spring 2013 and Fall 2013 semesters, and (ii) their standard deviations.

The values in Table 3.8 are drawn from 86 statements, whose wording was false in the Fall 2009 to Fall 2012 semesters and true in the Spring 2013 and Fall 2013 semesters (refer to Table 2.3). Inspection of the data in Table 3.8 shows again that the values for the Spring 2013 and Fall 2013 semesters are again numerically higher those

Semester	n	Mean Fraction Believed	Std. Dev.
Fall 2009	114	0.165	0.105
Fall 2010	107	0.136	0.100
Fall 2011	91	0.156	0.101
Fall 2012	91	0.093	0.095
Spring 2013	126	0.233	0.098
Fall 2013	110	0.234	0.088

Table 3.8. Mean fraction of incorrect beliefs endorsed per semester, using exclusively 86 statements (phrased as true for the Spring 2013 and Fall 2013 semester) of the original set of 215 statements

of the first four semesters (Fall 2009 to Fall 2012), which may account for the observed significance in misconception endorsement scores among the six semesters. To test whether or not the scores between the two semester sets are significantly different from each other, I performed an ANOVA test on the mean fraction of incorrect beliefs endorsed between the two sets. The variances in the mean fraction of incorrect beliefs endorsed between the sets, interestingly enough, were significantly different from each other ($W = 4.973$, $p_V = 0.026$). Nonetheless, an ANOVA test confirms that between the semester groups, the mean fractions of incorrect beliefs endorsed even after instruction once again differ significantly from each other ($F(1,637) = 130.4$, $p < 0.0005$, $\eta^2 = 0.170$, $\omega^2 = 0.168$), which leads one to make the same suggestions as in Sections 3.5.1 and 3.5.2, that the format of the ABI may play a significant role on the overall reported degree of misconception endorsement even after instruction in the course.

3.5.4 Correlation between true and false statements

As an additional test on the reliability of the ABI data, I measured the effect of any bias from the way in which statements were phrased. To do this, I correlated the mean fraction of misconceptions endorsed in the Spring 2013 and Fall 2013 semesters by preparing two special statement sets: one for the 129 incorrect statements; and one for the 86 correct statements. That is to say, I correlated the fraction of “false” statements endorsed with the fraction of "true" statements rejected, as outlined in Section 2.2.2 and mentioned on page 43. To run the correlation analysis, I created a data set with the student identifications in one column, the mean fraction of the 129 false statements believed in the second column, and the mean fraction of 86 true statements rejected in the third column. I then correlated the scores between the second and third columns. I found that the correlation between the endorsement of incorrect statements and the rejection of true statements is 0.361 and is significant ($r^2 = 0.14$, $df = 235$, $p < 0.0005$). That $r^2 = 0.14$ says that 14% of the variance in the mean fraction of misconceptions endorsed is accounted for by variation in the phrasing of the statements, as either true or false. So, *while there is a relationship between whether the misconceptions are endorsed due to statements being phrased one way vs. the other, the difference in phrasing accounts for only 14% of the variance, which suggests that the correlation is relatively weak.*

While the correlation measures the strength of the relationship between true and false statements, an additional test is necessary to discern whether or not there is a preference for students to score preferentially higher on statements phrased as either true (scientifically accurate) or false (incorrect). The data consists of the fraction of misconceptions endorsed by each student and is divided into two sets: one for false statements, and one for true statements. One can use a paired samples t-test (Walker, 1985, pp. 322-323) to compare the mean difference in the scores between the sets (here, fractions of misconceptions associated with true vs. false statements) with zero. The t statistic is established as follows: for a particular data analysis, scores x_i and y_i are associated with two variables representing data for subject i, for up to N pairs in the analysis under consideration. The formula for t is then given (p. 320) by

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2 - \frac{\left(\sum_{i=1}^{N}(x_i - y_i)\right)^2}{N}}{N(N-1)}}}, \tag{3.6}$$

where $\bar{x}$ and $\bar{y}$ are the respective means of the scores x and y for the N subjects in the analysis. The paired samples t-test also computes a p statistic, which determines if the differences in the means are *significantly* different from zero.

In continuing with my analysis of the effect of wording statements as true vs. false, I ran a paired samples t-test on the data from the Spring 2013 and Fall 2013 semesters, with x representing the fraction of false statements believed even after instruction and y representing the fraction of true statements rejected even after instruction. A paired samples t-test on the data reveals that there is a statistically significant preference for students to endorse misconceptions more so than they reject a true statement ($t = 5.77$, $p < 0.0005$). This result can be interpreted, for example, to mean that on a true-false-type questionnaire, students would be more likely to endorse the misconception that “Earth’s axis is not tilted compared to the ecliptic” more so than they would reject the fact that “Earth’s axis is tilted compared to the ecliptic.” The paired samples t-test illustrates that *there is a preference for students to endorse misconceptions more so than they reject scientifically-accurate statements*. This result is consistent with the notion that students who take an introductory-level course in astronomy have some tendency to believe what they hear, which suggests that *instructors should spend more time teaching in the context of common false beliefs, rather than simply focus on teaching fact by fact*. In Section 3.8, I provide additional support for this suggestion by testing for differences in the fractions of misconceptions endorsed between the Spring 2013 and Fall 2013 semesters.

3.5.5 Brief summary of results thus far

To summarize my results from Section 3.5 briefly, I compared the mean fraction of misconceptions endorsed from the Fall 2009 to Fall 2012 semesters with the mean fraction of misconceptions endorsed from the Spring 2013 to Fall 2013 semesters. Between these semesters, the format of the ABI was changed, in that 86 statements were rephrased as true. The effect of changing the format of the ABI due to the rephrasing of the 86 statements may have caused a significant increase in the overall fraction of misconceptions endorsed even after instruction in the course, because the results of Sections 3.5.1, 3.5.2, and 3.5.3 support that students in the Spring 2013 to Fall 2013 semesters report endorsing a significantly higher fraction of misconceptions than students in the Fall 2009 to Fall 2012 semesters. Furthermore, I observe that this increase does not depend on whether the associated statements are phrased as true or false, because the results of Section 3.5.4 support that responses between endorsement of misconceptions and rejection of correct statements are correlated.

My results seem to suggest that by rephrasing just under half of the statements (originally false) as true, students tend to score significantly worse on the ABI. The contrasting scores suggest that *student performance on the ABI may depend on the way that the statements are phrased and perhaps also on the instructions for completing the ABI*. In the first format of the ABI, students were told to indicate approximately when in their lives they realized that a statement was false, or if they still believe it, in a survey-like format. In the second format of the ABI, students were told to indicate approximately when in their lives they first realized that each statement is either true or false, in an exam-like format. At the onset, one might suppose that the discrepancy in the overall reported fraction of misconception endorsement may depend on the ABI format. In Section 5.2, I present additional results in support of this conjecture.

It turns out, however, that the change in the ABI format has relatively little effect on *correlations* between misconceptions. In Chapter 7, I explicitly outline a method to assess these correlations. Note that despite the variability in the overall fraction of misconceptions endorsed throughout all six semesters (Table 3.6), I show in Section 7.2.4 that such variability has no significant influence on the correlations between misconceptions. Additionally, in Section 7.2.4, I show that correlations between misconceptions are even less affected by rephrasing 86 of the 215 statements as true.

One caveat in my analysis of false vs. true statements remains, however, that in the Spring 2013 semester, students received instruction from DJB, as discussed in Section 2.1. What remains to be tested, then, is the effect to which students report a higher degree of misconception endorsement, even after instruction, with one professor (DJB) over another (NFC). Section 3.8 proceeds with a discussion of the effect of teaching pedagogy on one's own self-report of misconception endorsement.

3.6 Effect of statement wording

In the Fall 2009 to Fall 2011 semesters, students were encouraged to provide written feedback to the statements in the ABI which they thought were incorrect. The written feedback provides some quantitative assessment of the validity of the ABI as an instrument for assessing misconception endorsement. Namely, the feedback provides measures of:

1. the consistency between the misconception retainment scores (1, 2, 3) and the context of the written responses,
2. the consistency between the statement wording and its interpretation, and
3. whether or not the written feedback is an incorrect "correction" to the misconception (e.g., "the Sun isn't the brightest star, but Polaris is").

Rather than perform a comprehensive analysis of all 215 statements, I arbitrarily chose the following five statements of the ABI (refer to Appendix A): sA68, “we do not have telescopes in space,” sA172, “Saturn’s rings are solid,” sA189, “the Sun is the brightest star in universe,” sA226, “the galaxies are randomly distributed,” and sA263, “astronomical ideas of mass, distance, and temperature of planets are all speculative.” The overall mean degree of misconception retainment for the whole ABI, using data from all six semesters, is 1.88. For comparison, a score of 2 means that overall, students typically endorse misconceptions through adolescence, but not typically after instruction in the course. Hence, a score of 1.88 means that students less often endorse misconceptions through adolescence, and more often abandon misconceptions after instruction in the course. Statements sA68, sA172, sA189, score below the mean of 1.88, which makes them relatively “easy” items. On the other hand, statements sA226 and sA263 score above the mean, which makes them relatively “hard” items. In fact, of all 215 statements in the inventory, sA226 is associated with the fourth highest degree of misconception retainment. Therefore, these five statements tap into a range of misconceptions from “easy” to “hard.”

For this analysis, I arbitrarily chose to look at written responses from the Fall 2010 administration of the ABI. From the written feedback, I determined, for each statement, “% Wrong Code,” which is the percent of those *n* students whose written feedback is inconsistent with the response code (as determined by the bubble filled on the bubble sheet), “% Misinterpreted,” which is the percent of those *n* students whose written feedback indicates a misinterpretation of the statement itself, and “% Incorrect,” which is the percent of those *n* students whose written feedback is an incorrect statement. An example of a wrong code is if a student endorses a misconception but indicates a retainment score of “1” or “2.” An example of a misinterpreted statement would be if a student endorses that Saturn’s rings are solid, but then writes “transparent.” An example of incorrect feedback would be if a student rejects that the Sun is the brightest star in the universe, but then writes “Polaris is the brightest star.” Table 3.9 presents an examination of written feedback to each of the five statements, where *n* is the number of students who provided written feedback.

Statement	***n***	**% Wrong Code**	**% Misinterpreted**	**% Incorrect**
sA68	77	2.6	0.0	0.0
sA172	84	1.2	1.2	0.0
sA189	78	0.0	0.0	7.8
sA226	49	8.7	2.2	30.4
sA263	54	0.0	1.9	0.0

Table 3.9. Examination of written feedback to five statements of the ABI

Table 3.9 shows that students mistakenly fill in an incorrect bubble between 0% and 9% of the time, which indicates excellent agreement between the intended response and the actual response. Table 3.9 further shows that students misinterpret the selected statements only about 0% to 2% of the time. Hence, the frequency of either incorrect responses or statement misinterpretation at the end of the course is of relatively minor concern. For clarity, these results do not quantify the extent to which student recollections are reliable. A detailed analysis of recollection reliability is presented in Chapter 4.

Of the feedback that I had available to me, according to Table 3.9, I found that sA226, “the galaxies are randomly distributed,” is the most likely statement of the group to be associated with incorrect response codes (8.7%) and incorrect “corrections” (30.4%). The vast majority of the incorrect feedback consisted of statements about the galaxies being uniformly or evenly distributed. (In reality, the galaxies are distributed in clusters, voids, filaments, and knots.) My results for sA226 thus tentatively suggest that students are less likely to provide an accurate statement correction to the very hardest items in the ABI, compared to easier items.

3.7 Discussion

In this Chapter, I have performed a preliminary examination of the validity of the ABI as an instrument for assessing misconceptions held by students. From my examination, I have determined the following:

1. The presentation order of statements in the ABI has no significant influence on students' self-reports (Section 3.3).
2. The effect of fatigue in the process of completing the ABI has no significant influence on students' self-reports (Section 3.4). Hence, one need not be concerned with the length of the ABI.
3. There is a preference for students to endorse misconceptions more so than they reject scientifically-accurate statements (Section 3.5).
4. The change in the format of the ABI due to the rephrasing of the 86 items from false to true may cause a significant increase in the overall reported fraction of misconceptions endorsed even after instruction (Section 3.5, p. 51).
5. There do not seem to be significant issues with statement misinterpretation or incorrect response codes to selected statements (Section 3.6). However, there may be a higher tendency for students to provide incorrect feedback to only the very hardest items in the ABI.

Based on my findings thus far, depending on the particular format of the ABI that an instructor chooses to administer to students, the ABI may be a valid instrument for directly gathering information about misconceptions held by students. The validity of student self-reports, however, is also a reconstructive process based on memory recollection. I assess the consistency of student memories in Chapter 4. The extent to which correlations between misconceptions change from one semester to another is addressed in Chapter 7.

3.8 Teaching pedagogy

Thus far in this Chapter, I have established a method for scoring the tendency for misconceptions, each of which is associated with a statement in the ABI, to persist. I have also shown that the measured misconception retainment scores may vary significantly from one semester to another if only a subset of the students volunteer their time. Because I have data for the Spring 2013 and Fall 2013 semesters, in which two different instructors taught AST 109, I am also able to use the ABI data to investigate whether students retain more misconceptions after taking the course from one instructor vs. the other instructor. In this Section, I briefly look at the effect of teaching pedagogy on the tendency for misconceptions to persist.

Between the Spring 2013 and Fall 2013 semesters, two different teaching pedagogies were employed, as described in Section 2.1. The teaching pedagogy by DJB in the Spring 2013 semester consisted of conventional lecture accompanied by clicker questions, and brief homework problems were assigned to the students. The teaching pedagogy by NFC in the Fall 2013 semester consisted of conventional lecture, without clicker questions or homework problems, but with deliberate emphasis on misconceptions during instruction and in extra credit assignments.

To examine the influence of teaching pedagogy on the endorsement of incorrect statements or rejection of scientifically-accurate statements, I first performed an ANOVA test, using data from the 129 false statements, by comparing the fraction of incorrect statements believed between the Spring 2013 and Fall 2013 semesters. The variances in the mean fractions of misconceptions endorsed between the two semesters were not significantly different from each other ($W = 0.131$, $p_V = 0.717$). I found that the difference in the fraction of endorsed misconceptions associated with false statements is statistically significant ($F(1, 234) = 4.56$, $p = 0.034$, $\eta^2 = 0.019$, $\omega^2 = 0.015$). Because $p < 0.05$, this result is statistically significant and is consistent with the hypothesis that *addressing misconceptions is more effective in enabling* *students to reduce the number of misconceptions they endorse*. One implication is that an instructor may make a student more aware of their own misinformation more effectively by teaching why the misinformation is wrong. This result is consistent with my supposition in Section 3.5.4, that instruction ought to be guided with the awareness that students may endorse misconceptions more so than they reject scientifically-accurate statements.

The preceding analysis produced a statistically significant result ($p = 0.034$), though the effect size ($\eta^2 = 0.019$, $\omega^2 = 0.015$) is small, in the sense described in Section 3.2. To be clear, the small effect size does *not* imply that misconceptions cannot be dispelled in *small* classroom settings. Instead, the small effect size suggests that a large *sample* size is needed to obtain statistical significance in a repeated study. That is to say, the ABI was administered in two small classroom settings, each with a different instructor and pedagogy, but with all other aspects of the course remaining unchanged, then a statistically significant result may not be obtained, because the sample size may not be large enough.

Using data from the 86 scientifically-accurate statements, I then compared the fraction of misconceptions endorsed associated with these statements between the Spring 2013 and Fall 2013 semesters. The variances in the mean fractions of scientifically-accurate statements rejected between the two semesters were not significantly different from each other ($W = 2.127$, $p_V = 0.146$). I found that the difference in the fraction of endorsed misconceptions associated with scientifically-accurate statements is not statistically significant ($F(1, 234) = 0.001$, $p = 0.98$, $\eta^2 = 0.000$, $\omega^2 < 0$). This result indicates that there is no evidence to suggest that either teaching pedagogy is necessarily better than the other at helping students to learn the *correct* information.

To summarize, instruction targeting student misconceptions may have helped students to reject statements that seem false, because the difference in the fraction of endorsed misconceptions associated with *incorrect* statements is significantly lower for students taught by NFC than by DJB. Also, instruction targeting student misconceptions does not seem to help students to endorse truthful statements, because there is no significant difference in the fraction of endorsed misconceptions associated with true statements. Ultimately, my analysis supports that with regard to overall misconception endorsement, *there is a marginal benefit in teaching students in the context of their misconceptions over teaching students using conventional fact-oriented lecture*.

CHAPTER 4

RELIABILITY OF STUDENT RECOLLECTION OF THEIR PAST BELIEFS

The ABI provides a direct measure of the degree to which students will endorse or reject misconceptions either before or after instruction in the college classroom. Student self-reports, however, involve a reconstructive process based on their ability to recall memories of when they disabused themselves of misconceptions. While the ABI may well serve as a post-instruction examination of student misconceptions, pre-instruction examinations are also necessary to ensure the accuracy of student self-reports of their past beliefs prior to instruction. I considered two influences on the extent of one's own recollection reliability: (i) the student's self-reported childhood interest in astronomy, prior to instruction in the course; and (ii) the student's self-reported level of stress in the semester. Motivated by the results of my analysis of the ABI through the end of the Spring 2013 semester, I designed two instruments for the Fall 2013 semester to assess childhood interest in astronomy and stress. In this Chapter, I examine the reliability of student recollection based on their self-reports to statements in the ABI.

4.1 New instruments to measure childhood interest in astronomy and stress

One might expect that those students who took an early-in-life interest in astronomy are more likely to recall their memories of what they learned. On the other hand, memory recall is a reconstructive process, so one might expect that high levels of stress in one's life may interfere with the reconstructive process, thus causing inconsistencies in one's own past recall of their own misconceptions (Kuhlmann, Piel, & Wolf, 2005). In response to these considerations, I developed an instrument to probe childhood interest in astronomy, at the beginning of the course, and another instrument to probe one's stress level, at the end of the course.

4.1.1 Childhood interest in astronomy

The Childhood Interest in Astronomy Survey (CIAS), designed together by the dissertation author and NFC, is a survey of eight questions which asks students to indicate their childhood interest in various activities related to astronomy. All students taking AST 109 in the Fall 2013 semester were reminded to complete the CIAS, on the University of Maine Learning Management System, by the end of the second week of the course for one point extra credit on their final grade. Students were denied access to the CIAS after the second week. The implementation of the CIAS to the overall research project was formally approved by the IRB on August 21, 2013.

A total of 80 students completed the CIAS. The questions and response options are presented in Appendix C. For each question, students were asked to report their interest level for each activity, with the options for each question being: "never," "occasionally," and "very often." For example, in Question 1 of the CIAS, if a student has read astronomy books very often during one's own free time, the student responded to Question 1 with "very often." To quantify the responses on a kind of Likert metric, I developed a master code for the responses to the CIAS, as indicated in Table 4.1. These options are ranked in order from low to high childhood interest. For each student, I calculated the overall childhood interest in astronomy by summing the scores for all of the eight questions of the CIAS. (Only one student skipped one question, which was scored as 0 and added to the scores of the remaining seven questions.) Scores fell within the range of 0 to 11, with a maximum possible score of 16. The mean and standard deviation of scores on the CIAS were 4.57 and 2.56, respectively, which suggests that the majority of students who took the CIAS had little more than an occasional interest in activities related to astronomy as a child.

Never	Occasionally	Very Often
0	1	2

Table 4.1. Codes for the CIAS

One concern about the CIAS is that telescopes, planetaria, and observatories may not have been available to some students, who may have been genuinely interested in astronomy as children. For example, one's interest in astronomy may be influenced if a student has access to these instruments while still a child. Because the CIAS is a new instrument in itself, I elected to test the internal consistency of the survey, using coefficient alpha (refer to Section 3.4) as the test statistic. Values of $\alpha \geq 0.70$ represent a group of items with "adequate" internal consistency (Schmitt, 1996), which is satisfactory enough for the purpose of correlating the responses with data from other instruments. I found that for the CIAS, $\alpha = 0.72$. The internal consistency of the responses would decrease if any one of the eight questions was removed from the survey, except for Question 3, which asks about how often the student used binoculars or a telescope, in which case α increases only marginally to 0.73. My results support that I can use data from all eight questions without sacrificing the reliability of the data.

With the responses to all eight questions from the CIAS available, I then compared the childhood interest in astronomy scores for all students with their mean misconception retainment scores for the Fall 2013 semester, as determined by their ABI responses and calculated in the manner described in Section 3.1. Correlation statistics are presented here. The correlation between childhood interest in astronomy and mean misconception retainment score is -0.470 ($r^2 = 0.22$, $df = 55$, $p < 0.0005$), where r^2 is described in Section 3.5.4. The result presented here states that 22% of the variance in childhood interest in astronomy is accounted for by variation in the mean misconception retainment score with the variables being anti-correlated. The result is statistically significant and suggests preliminarily that *the more interested children are in astronomy, the less likely they are to retain misconceptions*. Another interpretation is provided in the subsequent paragraph. Figure 4.1 plots childhood interest in astronomy for students in the Fall 2013 semester vs. their mean misconception retainment scores.

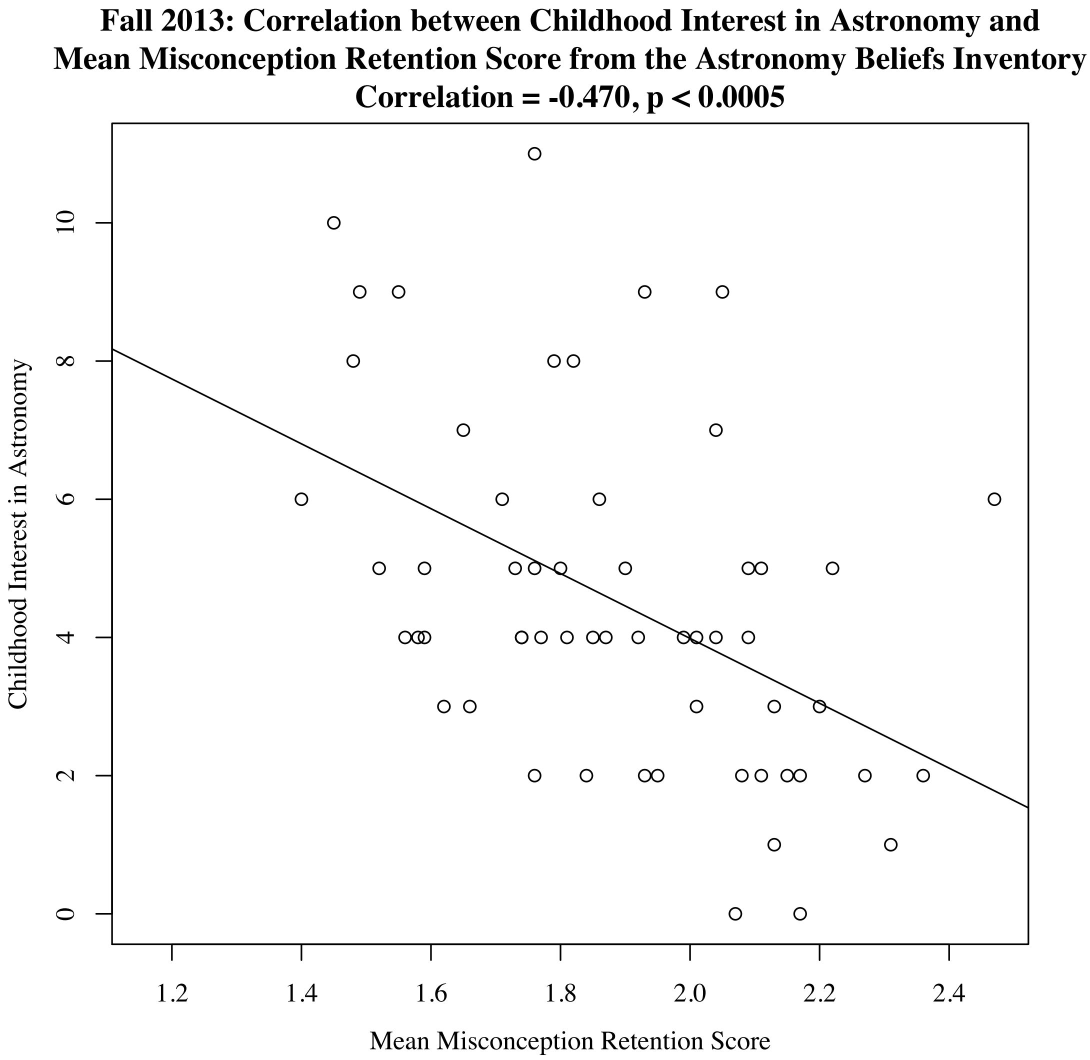


Figure 4.1. Childhood interest in astronomy vs. mean misconception retainment score for students in the Fall 2013 semester

The scale for childhood interest in astronomy is the sum of the scores for all eight questions of the CIAS. For the plots in Figure 4.1 and the remaining figures in Chapter 4 a least squares fit line has been included if the significance $p < 0.05$.

To check the extent that the CIAS can consistently probe the dispelling of misconceptions as a child or adolescent, I recoded the ABI data into two categories: one for endorsing the incorrect belief through adolescence, and one for disabusing oneself of the incorrect belief during childhood or adolescence. Table 4.2 presents a summary of the recoded responses, where the code “0” indicates that one disabused oneself of a misconception during childhood or adolescence, and the code “1” indicates that a misconception was retained all the way through adolescence.

Degree of Misconception Retainment	Retained Misconception
1 →	→ 0
2 →	→ 1
3 →	→ 1

Table 4.2. Recoded responses indicating retention of a misconception through adolescence

Using the recoded data, I compared the childhood interest in astronomy scores for all students with the mean fraction of misconceptions endorsed through their adolescences. The correlation between childhood interest in astronomy and mean fraction of misconceptions endorsed through adolescence is −0.495 ($r^2 = 0.25$, $df = 55$, $p < 0.0005$). That childhood interest in astronomy accounts for 25% of the variance in misconception retention (with the variables again being anti-correlated) further supports that the more interested children are in astronomy, the less likely they are to endorse misconceptions after college instruction.

4.1.2 Stress

Originally designed by S. Cohen, Kamarck, and Mermelstein (1983), the Perceived Stress Scale is a 14-question survey which asks students to report on their own measure of stress in general. For each question in the survey, subjects indicate the extent to which a particular facet represents them (e.g., how often one has felt overwhelmed). Scores for each question range from 0 to 4, where 0 means the subject has never experienced the facet, and 4 means the subject has very often experienced the facet. A related survey designed to assess one's own stress would be most apropos to administer simultaneously with the administration of the ABI.

Given the large number of statements in the ABI, however, thesis co-advisor Prof. Geoff L. Thorpe (GLT) and I agreed to administer a shortened version of the Perceived Stress Scale to minimize the bias in one's own stress self-report. From the original 14-question instrument, we created a shortened survey, consisting of four questions which were determined by GLT as being most inter-correlated and comprising a sufficiently large proportion of the variance (private communication). In statistics, one is often interested in retaining data that account for as much of the variance as possible, because the higher the variance accounted for, the more likely subjects respond with all of the available scores (0 to 4), which provides a better statistical representation about the distribution of scores. In contrast, data that account for very little variance tend to be restricted to only a couple of scores (e.g., 0 to 2), because subjects rarely respond with the other available scores.

The set of these four questions comprises the Stress Survey. All students taking the ABI were first asked to complete the stress survey prior to starting the ABI. All students who completed the survey received one point extra credit on their final grade. On the informed consent, students were provided with information, if the student deemed necessary, to consult a counselor on the University of Maine campus free of charge. The implementation of the Stress Survey to the overall research project was formally approved by the IRB on August 21, 2013.

A total of 108 students completed the Stress Survey. The questions, response options, and procedure of the Stress Survey, as presented to the students, are presented in Appendix D. In Questions 1 and 4, higher scores indicate higher stress. In Questions

2 and 3, the metric is reversed, so that lower scores indicate higher stress. To quantify the responses on a kind of Likert metric, I considered the direction of the scores for all four questions, and ultimately I preserved the codes as originally presented to the students, summarized in Table 4.3. For each student, I calculated the overall stress by summing the scores for all of the four questions of the Stress Survey, and by taking into consideration the reversed direction of the metric for Questions 2 and 3 of the Stress Survey. The total stress score per student is 8 + Q1 + Q4 – Q2 – Q3, where Q1, Q2, Q3, and Q4 are the scores for each of questions 1-4, respectively. (Note that the number of questions has nothing to do with the metric of the response codes.) Scores fell within the range of 0 to 16. The mean and standard deviation of scores on the Stress Survey were 6.63 and 3.47, respectively, indicating that at the time of administration of the ABI (only days before the final exam), the majority of students who took the Stress Survey reported moderate stress in their lives.

Never	Almost Never	Sometimes	Fairly Often	Very Often
0	1	2	3	4

Table 4.3. Codes for the Stress Survey

4.1.3 Relationship between childhood interest in astronomy and stress

In order to determine whether or not the CIAS and Stress Survey are independent instruments, I then compared the stress scores for all students with their mean misconception retainment scores for the Fall 2013 semester, as determined by their ABI responses and calculated in the manner described in Section 3.1. The correlation between stress and mean misconception retainment score is 0.049 ($r^2 < 0.01$, $df = 107$, $p = 0.61$), which indicates that less than 1% of the variance in stress is accounted for by the variance in mean misconception retainment, indicating that *stress has no significant influence on misconception retainment*. Figure 4.2 plots stress for students in the Fall 2013 semester vs. their mean misconception retainment

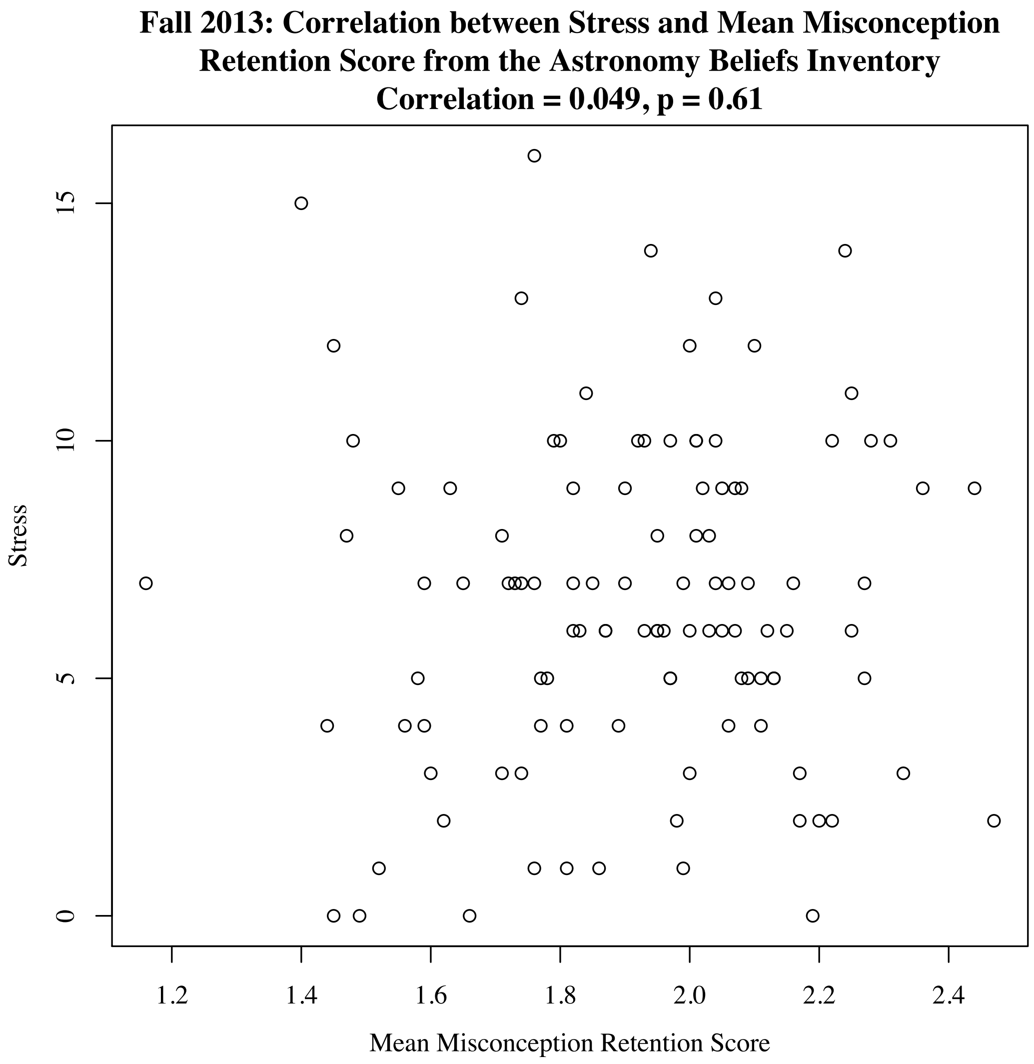


Figure 4.2. Stress vs. mean misconception retainment score for students in the Fall 2013 semester

scores. Additionally, the correlation between the data from the CIAS and the Stress Survey is -0.092 ($r^2 < 0.01$, $df = 55$, $p = 0.49$), which represents that less than 1% of the variance in stress is accounted for by the variance in childhood interest in astronomy, indicating that the CIAS and Stress Survey can be treated as independent instruments.

4.2 Black holes and galaxies pretest

4.2.1 Description

After the Fall 2011 semester, I had sufficient data from all administrations of the ABI to test a series of predictions regarding the order to teach black holes and galaxies, as outlined in Chapter 14. Motivated by my results, I worked with DJB to design video lectures for each topic, as well as pretests and posttests associated with each video lecture, to assess whether or not the order of instruction significantly affects how much students learn from lecture. The design of the study is presented in Chapter 16.

It turns out that the questions on the pretests associated with the video lectures also provide a useful measure from which I can assess incorrect beliefs held by the students, because the questions and response options are specifically tailored to the statements on the ABI and incorrect beliefs held by the students. The wording of the questions and response options, for both the black hole and galaxy topics, went through multiple revisions by NFC, GLT, and the dissertation author, after which we developed 18 multiple-choice questions with incorrect answers designed and motivated by *a priori* knowledge of student misconceptions. The incorrect answers serve as “distractors,” in the sense that they distract students from endorsing the correct information by prompting memory recall of their prior misconceptions (Sadler, 1998). The 18 questions and response options of the instrument, termed the Black Holes and Galaxies Pretest (BHGP), are presented in Appendix E. Table 4.4 presents the BHGP

ABI Statement(s)	% Correct	Question
sA232	40%	1. How are black holes created today?
sA233, sA245, sA247	30%	2. What is the fate of black holes?
sA235, sA237	73%	3. Black holes consist of:
sA246	14%	4. As detectable from our universe, what shape does a black hole have?
sA238	9%	5. A spacecraft is put into circular orbit around a distant black hole. Which one of the following outcomes will happen to the spacecraft?
sA240	67%	6. What would happen to an asteroid that passed into a black hole?
sA242	70%	7. How do astronomers detect black holes?
sA243, sA244	65%	8. If we boarded a spaceship to journey into a black hole, what would happen to us?
sA218, sA225	74%	9. How many galaxies exist in the universe today?
sA219	12%	10. Relative to the Milky Way galaxy, the solar system is located:
sA220	42%	11. How many distinct shapes do galaxies have?
sA221, sA224	53%	12. Where in the universe is the Milky Way located today?
sA222	18%	13. Where is the Sun located relative to the Milky Way?
sA226	20%	14. How are galaxies distributed throughout the universe?
sA227	19%	15. Which statement about observing stars in the Milky Way is most accurate?
sA228	80%	16. Which one statement about galaxy properties is correct?
sA230	82%	17. Which statement most accurately describes the contents of the Milky Way?
sA231	34%	18. Which one of the following is true about the Milky Way?

Table 4.4. Black Holes and Galaxies Pretest questions, the percent of students who answered correctly, and the associated statement(s) in the ABI

questions, the percent of students who answered the question correctly, and each of the associated ABI statement(s) for each question.

Administration of the BHGP in the Fall 2013 semester proceeded as follows. Students in AST 109 were told that they would complete a two-part questionnaire. Part 1 consisted of the BHGP and was completed immediately after the students took the second prelim (a “prelim” is more commonly known in other university settings as “tests” or “exams”). Part 2 consisted of the ABI and was administered at the end of the semester. The BHGP was administered before any instruction was given on either black holes or galaxies. Students were explicitly reminded to complete both parts in order to receive a total of four points extra credit to their final grade. In the Fall 2013 semester, students in AST 109 were permitted 60 minutes to complete the second prelim, after which students would report to a testing room elsewhere in the building to complete the BHGP. On average, students took about 10 minutes to complete the BHGP. The BHGP received final approval from the IRB on October 29, 2013. The second prelim in the course, along with the BHGP, was administered on October 31, 2013. Scores on the BHGP fell within the range of 11.1% to 77.8%. The mean and standard deviation of scores on the BHGP were 44.4% and 13.3%, respectively.

Often the notion of a “pretest” prompts the expectation of a “posttest” often immediately following the associated lecture. For comparison, in the Fall 2013, the ABI served as the “posttest.” In the Fall 2013 semester, the ABI was administered to students who selected one of several available time slots after the last lecture and prior to the final exam. The time gap between administrations of the BHGP and the ABI is approximately one and a half months. Hence, comparisons between the BHGP and the ABI provide indications of *long-term consistency* between what students think they know after instruction and what students thought they knew prior to instruction.

Because the BHGP was administered prior to instruction on either black holes or galaxies in the course, responses to the BHGP illuminate whether or not students endorsed the correct information related to black holes. Student responses to the BHGP can be compared against their responses to the associated ABI statement(s). In total, I made 24 comparisons. Note that for each of the black hole questions, students were permitted to indicate if they denied the existence of black holes altogether. Because only two students of the 148 had indicated on the BHGP that they do not believe that black holes exist, I cannot provide a meaningful statistical analysis that compares the students' pre-instructional beliefs on the existence of black holes with their responses to the associated ABI statement, sA234, "black holes really don't exist." Therefore, I have decided not to compare any of the BHGP responses with responses to sA234.

Using the BHGP and ABI, in combination, I created three categories: "definitely consistent recollection," "definitely *in*consistent recollection," and "cannot conclude either way." Definitely consistent recollections are those in which either a student answers the pretest question correctly and reports having dispelled the associated misconception as a child or adolescent (misconception retainment score of 1), or the student answers the pretest question incorrectly, and still endorses the associated misconception (misconception retainment score of 3). Definitely *in*consistent recollections are those in which a student answers the pretest question one way, and reports the exact opposite degree of misconception retainment for the associated statement on the ABI. For those statements on the ABI in which a student does not report either pre-instruction disambiguation or post-instruction endorsement of a misconception (that is, if the misconception retainment score is 2), I cannot establish *definitely* whether or not a student's past recollection is consistent. Since each question on the pretest is associated with one or more statements in the ABI, I can use both instruments to determine the appropriate category of memory recollection. Figure 4.3 presents a flow chart illustrating how I determine whether or not student recalls of their own past memories are consistent. Respectively, the overall percentages of

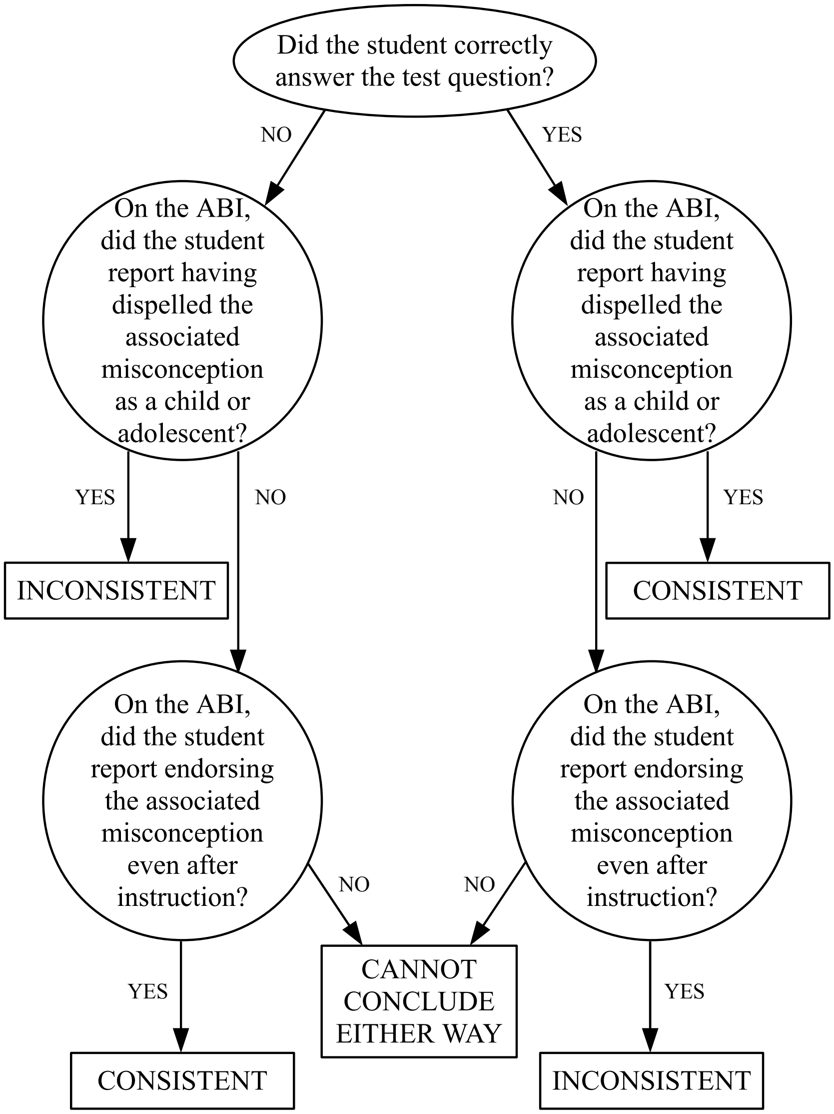


Figure 4.3. Flow chart to determine if student recollection of their pre-instructional beliefs on black hole and galaxy topics are consistent

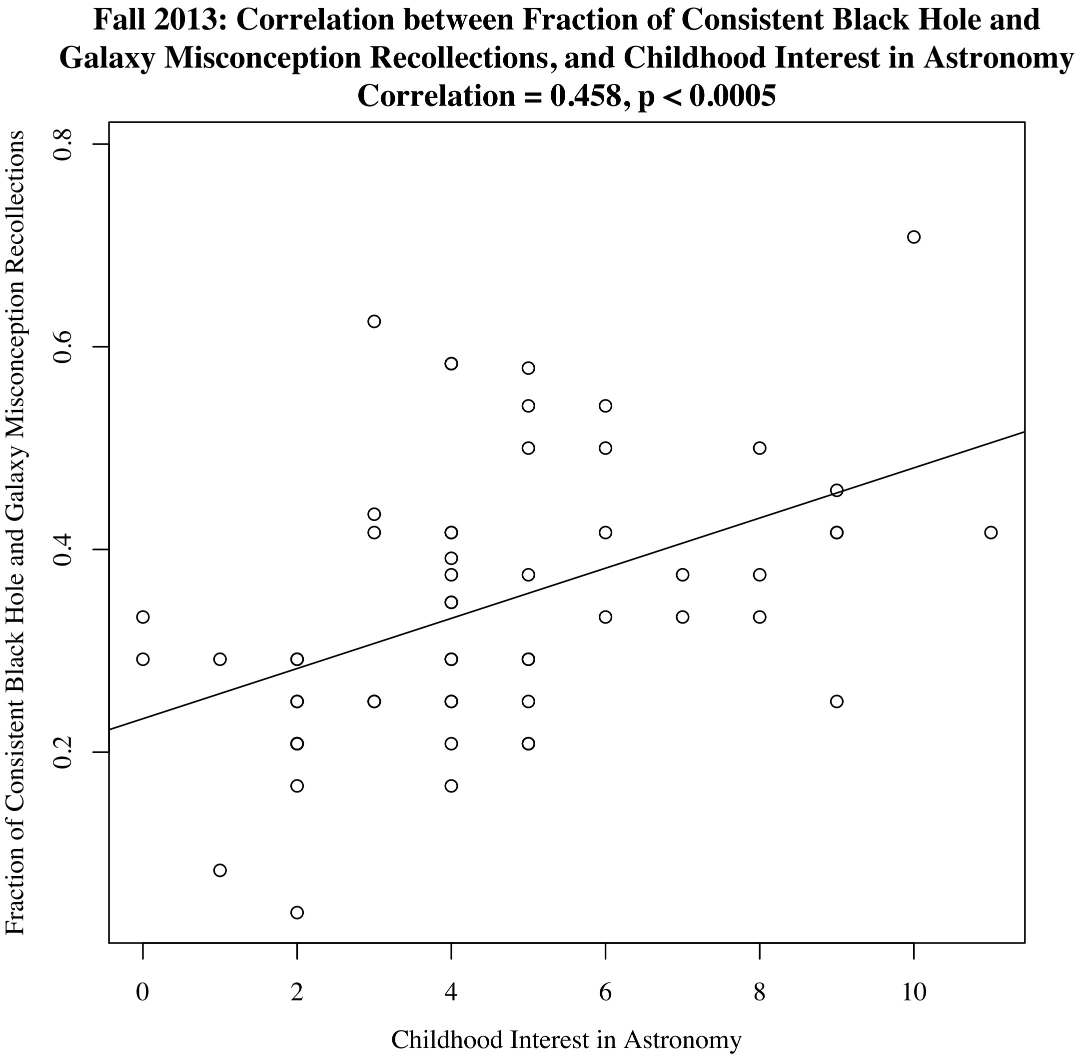
definitely consistent recollections and definitely *in*consistent recollections are 34% and 20%.

4.2.2 Black hole and galaxy recollection consistency vs. childhood interest in astronomy

Using the flow chart in Figure 4.3 and the 24 comparisons between BHGP questions and associated ABI statements, presented in Table 4.4, for each student, I determined the fraction of both consistent and inconsistent misconception recollections, as well as the fraction of the time in which I cannot tell either way. I then compared the fraction of both consistent and inconsistent black hole and galaxy misconception recollections with their childhood interest in astronomy.

First, I compared the fraction of consistent recollections with their childhood interest in astronomy. I found that the correlation between the fraction of consistent recollections and childhood interest in astronomy is 0.458 ($r^2 = 0.21$, $df = 55$, $p < 0.0005$), which represents that 21% of the variance in consistent black hole and galaxy misconception recollections is accounted for by the variance in mean misconception childhood interest in astronomy. The result indicates tentatively that increased childhood interest in astronomy is correlated with increased accurate recollection of black hole and galaxy misconceptions. Figure 4.4 presents a plot of the fraction of consistent black hole and galaxy misconception recollections vs. childhood interest in astronomy.

Then I compared the fraction of *in*consistent recollections with their childhood interest in astronomy. I found that the correlation between the fraction of *in*consistent recollections and childhood interest in astronomy is 0.320 ($r^2 = 0.10$, $df = 55$, $p = 0.015$), which represents that 10% of the variance in *in*consistent black hole and galaxy misconception recollections is accounted for by the variance in mean misconception childhood interest in astronomy. The result indicates tentatively that



Fall 2013: Correlation between Fraction of Consistent Black Hole and Galaxy Misconception Recollections, and Childhood Interest in Astronomy
Correlation = 0.458, p < 0.0005

Figure 4.4. Fraction of consistent black hole and galaxy misconception recollections vs. childhood interest in astronomy

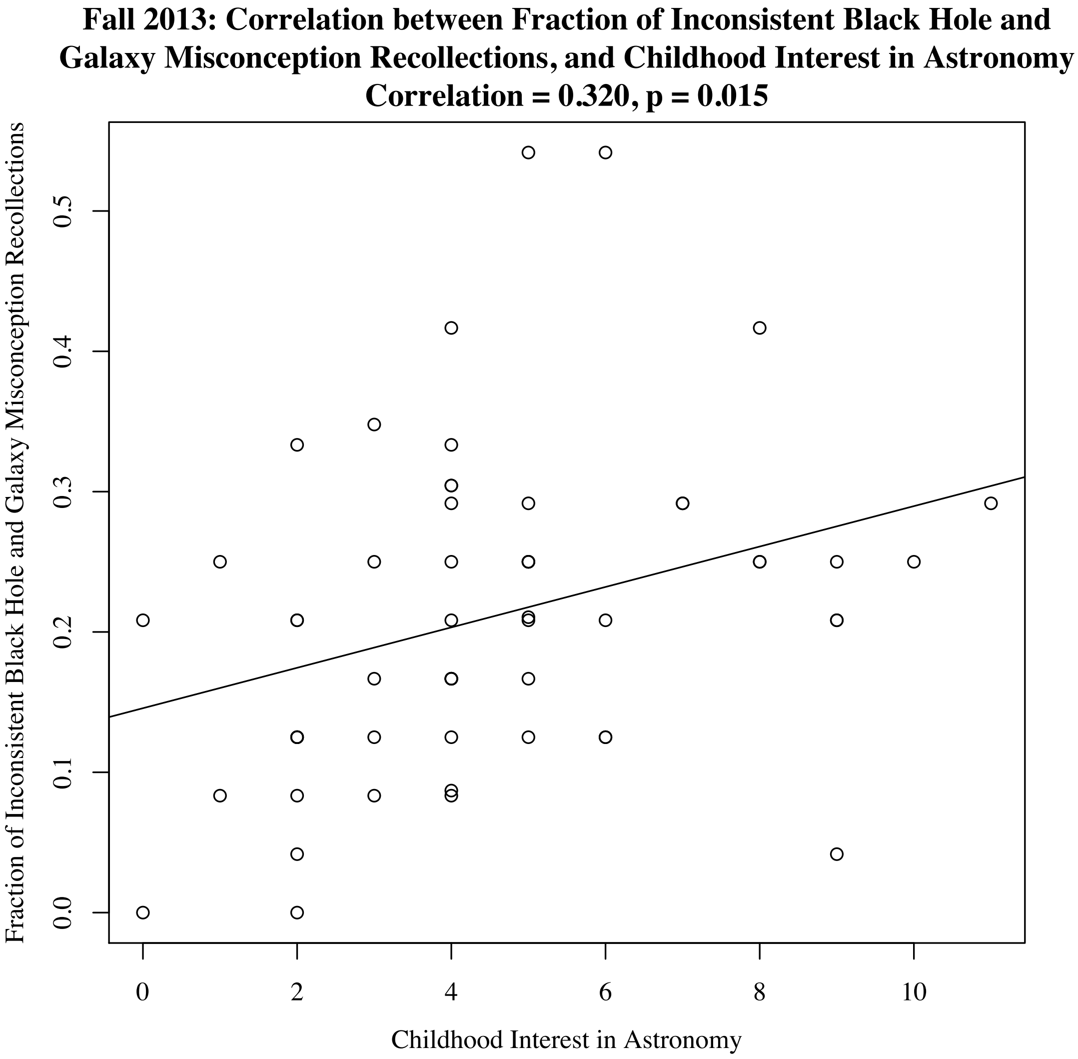


Figure 4.5. Fraction of inconsistent black hole and galaxy misconception recollections vs. childhood interest in astronomy

increased childhood interest in astronomy is correlated with increased *in*accurate recollection of black hole and galaxy misconceptions. Figure 4.5 presents a plot of the fraction of consistent black hole and galaxy misconception recollections vs. childhood interest in astronomy.

The previous results regarding correlations of childhood interest in astronomy and the consistency of memory recollections seem contradictory. How can increased childhood interest in astronomy lead to both increased recollection consistency and increased recollection *in*consistency at the same time? The answer lies in the "in between" category, in which I am unable to determine *definitely* whether or not one's own recollections are consistent. In the first analysis, I showed that increased childhood interest in astronomy led to a significant increase in definite recollection consistency. In the second analysis, I showed that increased childhood interest in astronomy led to a marginal increase in definite recollection *in*consistency. However, I do not have data to support that students with a *low* childhood interest in astronomy are definitely recalling their own past beliefs either consistently or *in*consistently. Together, both analyses support that (i) memory recollections from those students who indicate relatively *low* childhood interest in astronomy cannot be categorized as being definitely consistent or definitely *in*consistent overall, whereas (ii) memory recollections from those students who indicate relatively *high* childhood interest in astronomy *can* be categorized as being definitely consistent or definitely *in*consistent overall. In other words, *increased childhood interest in astronomy may be associated with a greater ability to discern whether or not one's memories are indeed accurate.*

4.2.3 Black hole and galaxy recollection consistency vs. stress

I then compared the fraction of both consistent and inconsistent black hole and galaxy misconception recollections with the self-reported stress level of the students. First, I compared the fraction of consistent recollections with their stress. I found that the correlation between the fraction of consistent recollections and stress is -0.154 ($r^2 = 0.02$, $df = 105$, $p = 0.11$), which represents that only 2% of the variance in consistent black hole and galaxy misconception recollection is accounted for by the variance in stress. The result indicates tentatively that there is no relationship between the fractions of consistent recollections and one's stress level, as measured. Figure 4.6 presents a plot of the fraction of consistent black hole and galaxy misconception recollections vs. stress.

I then compared the fraction of *in*consistent black hole and galaxy misconception recollections with the self-reported stress level of the students. First, I compared the fraction of consistent recollections with their stress. I found that the correlation between the fraction of *in*consistent recollections and stress is 0.133 ($r^2 = 0.01$, $df = 105$, $p = 0.17$), which represents that only 1% of the variance in consistent black hole and galaxy misconception recollection is accounted for by the variance in stress. The result indicates tentatively that increased stress also does not lead to any significant increase in recollection *in*consistency. Figure 4.7 presents a plot of the fraction of consistent black hole and galaxy misconception recollections vs. stress. Hence, the data from the BHGP and ABI together suggest that *stress does not significantly influence memory recollection consistency*.

4.3 Course prelims

4.3.1 Description

Prior to administration of the ABI in the Fall 2013 semester, NFC administered three prelims, each containing 50 multiple-choice questions. The prelims were not cumulative; each prelim tested material that was recently taught in class. As an additional check on the consistency of misconception recollections, I compared each ABI statement with each associated question from the three prelims, and I assembled a master list of the most closely-related ABI statement and prelim question pairs. For

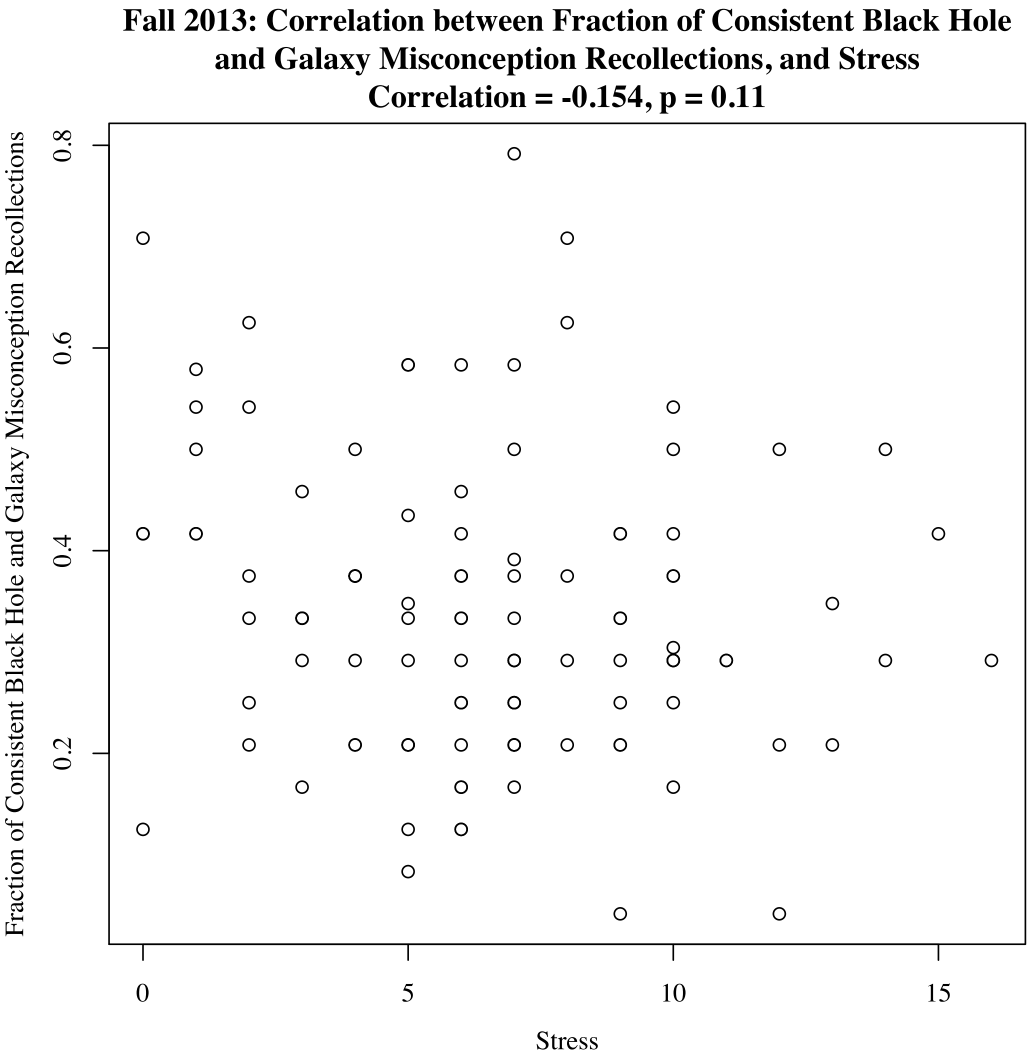


Figure 4.6. Fraction of consistent black hole and galaxy misconception recollections vs. stress

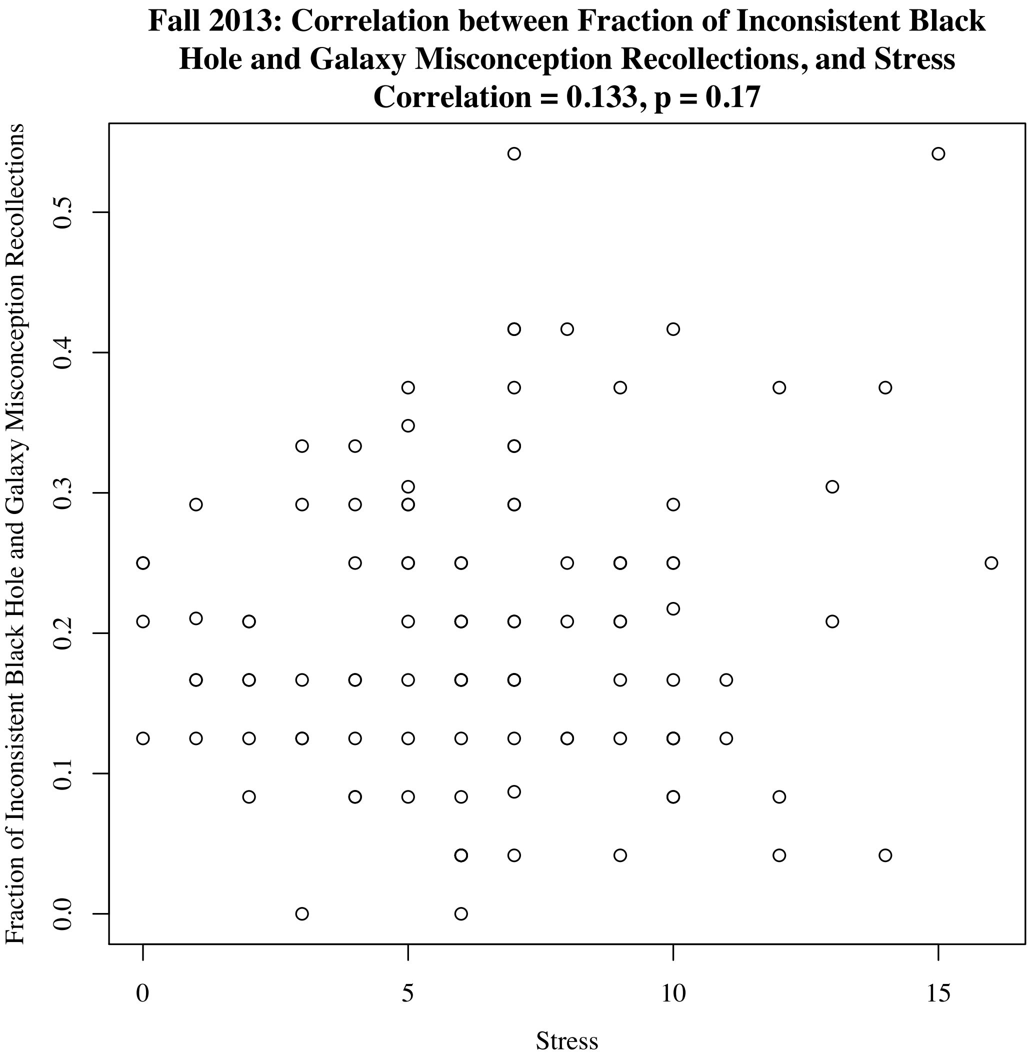


Figure 4.7. Fraction of inconsistent black hole and galaxy misconception recollections vs. stress

each prelim question, I scored all the student responses as correct or incorrect, by comparing their responses against an answer key for the prelim. In total, I made 14 comparisons between prelim questions and ABI statements. The prelim questions are scored as “correct” and “incorrect.” Responses to the ABI statements are scored as degrees of misconception retainment as presented in Table 2.4. Appendix F presents the list of ABI statements and their associated prelim questions.

Using the three prelims and ABI, in combination, I created two categories: “consistent recollection,” and “*in*consistent recollection.” The process of comparing prelim question responses to ABI statement responses has a particular advantage over the process that I used for comparing BHGP responses (described in Section 4.2), since the prelim questions are administered post instruction. The advantage is that I can use all three degrees of misconception retainment to discern, more definitely, whether or not one’s own recollection is consistent throughout the course. Since each relevant question on the prelims is associated with one or more statements in the ABI, I can use both instruments to determine the appropriate category of memory recollection. Figure 4.8 presents a flow chart illustrating how I determine whether or not student recalls of their own past memories are consistent. The overall fraction of consistent responses between the selected ABI statements and their associated prelim questions is 68%, which still leaves 32% of the responses inconsistent between the two instruments.

4.3.2 Prelim and ABI recollection consistency vs. childhood interest in astronomy

Using the flow chart in Figure 4.8 and the 14 comparisons between prelim questions and associated ABI statements, presented in Appendix F, for each student, I determined the fraction of consistent misconception recollections. I then compared the fraction of consistent recollections with their childhood interest in astronomy. I found that the correlation between the fraction of consistent recollections and childhood

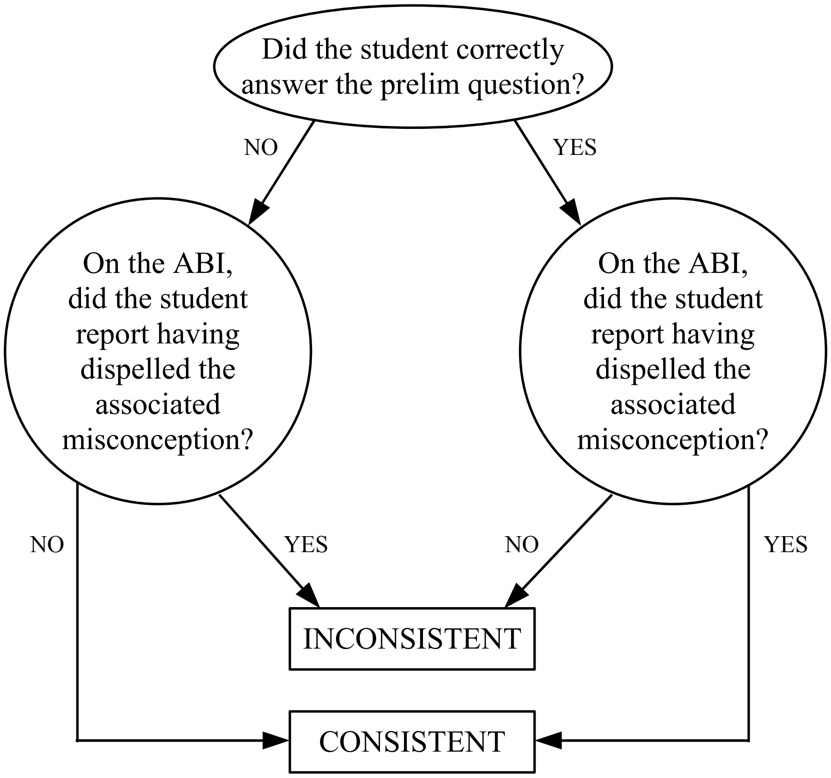


Figure 4.8. Flow chart to determine if student recollection of their own misconceptions are consistent with their prelim responses

interest in astronomy is 0.188 ($r^2 = 0.04$, $df = 55$, $p = 0.16$), which represents that only 4% of the variance in consistent misconception recollection is accounted for by the variance in childhood interest in astronomy. The result indicates that there is no association between increased childhood interest in astronomy and improved recollection consistency of one's own past misconceptions. Figure 4.9 presents a plot of the fraction of consistent misconception recollections, using prelim data, vs. childhood interest in astronomy.

4.3.3 Prelim and ABI recollection consistency vs. stress

I then compared the fraction of consistent recollections with the self-reported stress level of the students. I found that the correlation between the fraction of consistent recollections and stress is −0.068 ($r^2 < 0.01$, $df = 107$, $p = 0.48$), which

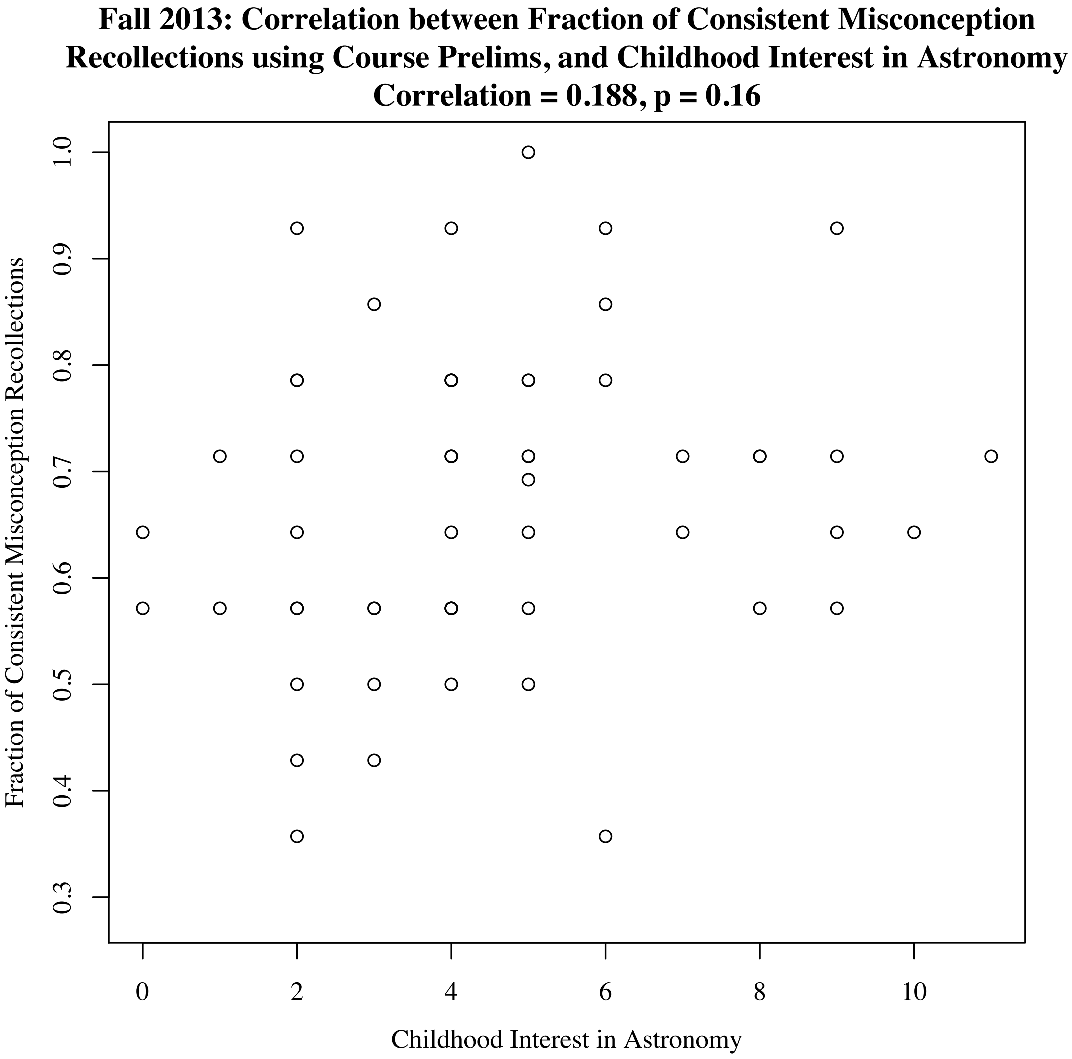


Figure 4.9. Fraction of consistent misconception recollections, using prelim data, vs. childhood interest in astronomy

represents that less than 1% of the variance in consistent misconception recollection is accounted for by the variance in stress. The result indicates that stress has *no significant influence* on recollection consistency. This result is consistent with that in Section 4.2.3. Figure 4.10 presents a plot of the fraction of consistent misconception recollections, using prelim data, vs. stress.

4.4 Unlearning misconceptions vs. endorsing the correct science

4.4.1 Description

After administration of the ABI in the Fall 2013 semester, NFC administered a cumulative final exam containing 100 multiple choice questions. As an additional check on the consistency of misconception recollections, I compared each ABI statement with each associated question from the final exam, and I assembled a master list of the most closely-related ABI statement and final exam question pairs. For each final exam question, I scored all the student responses as correct or incorrect, by comparing their responses against an answer key for the final exam. In total, I made 20 comparisons between final exam questions and ABI statements. The final exam questions are scored as “correct” and “incorrect.” Responses to the ABI statements are scored as degrees of misconception retainment as presented in Table 2.4. Appendix G presents the list of ABI statements and their associated final exam questions.

Because the final exam came *after* the ABI, however, I cannot necessarily discern the consistency of one’s own past recollections. On the other hand, however, I can use the two instruments together to determine the extent to which the disambiguation of a specific misconception may trace the endorsement of the correct associated science. Hence, using the final exam and ABI, in combination, I created two categories: “consistent,” and “*in*consistent,” where a consistent result is one in which either (i) the disambiguation of a misconception is associated with learning the correct science or (ii) the retainment of a misconception is not associated with learning the

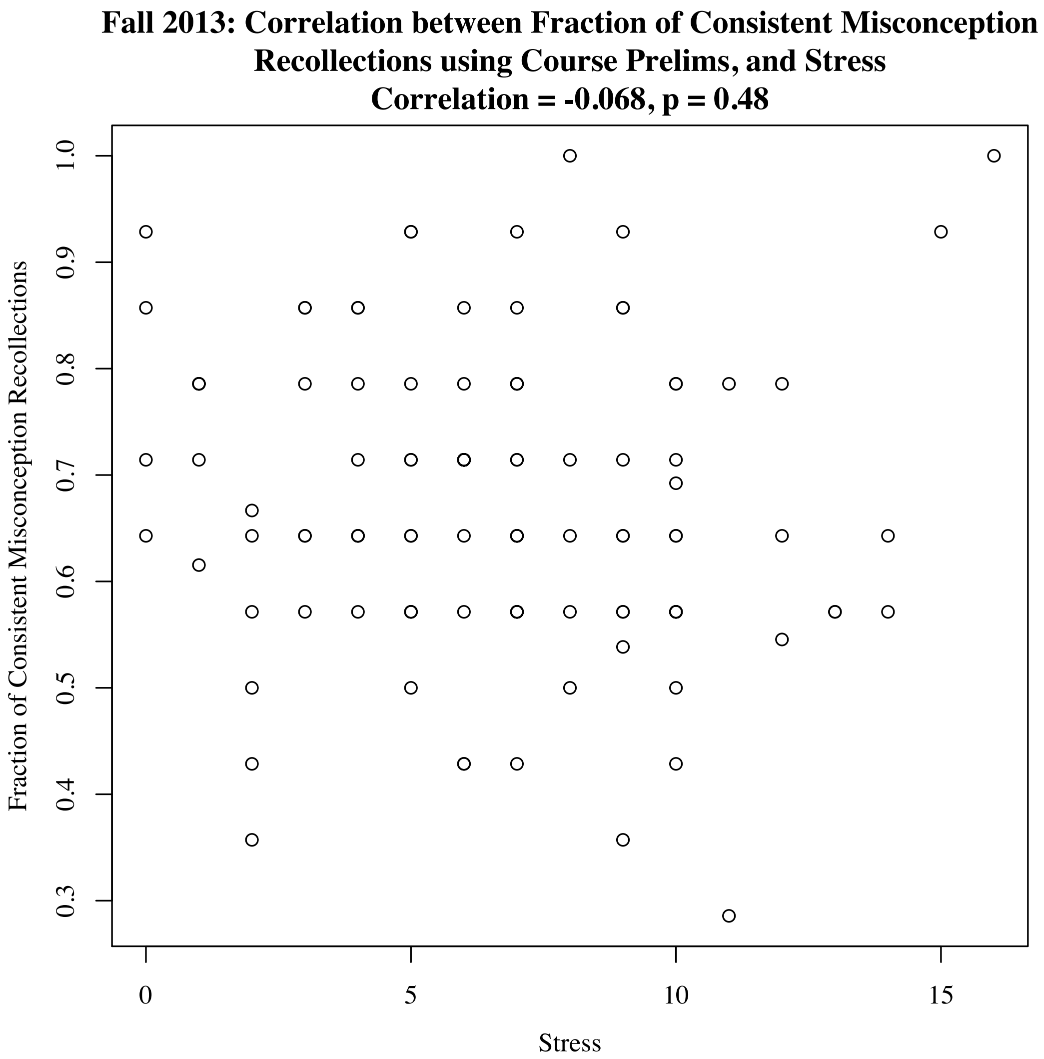


Figure 4.10. Fraction of consistent misconception recollections, using prelim data, vs. stress

correct science. Since each relevant question on the final exam is associated with one or more statements in the ABI, I can use both instruments to determine the appropriate category. Figure 4.11 presents a flow chart illustrating how I determine whether or not student recalls of their own past memories are consistent.

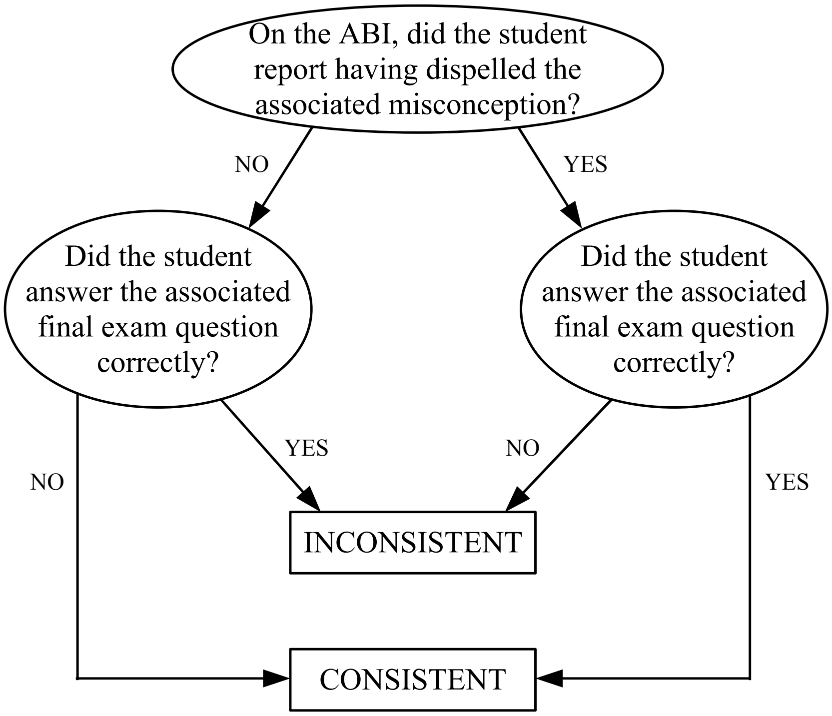


Figure 4.11. Flow chart to determine the consistency of responses between statements in the ABI and their associated final exam questions

The overall fraction of consistent responses between the selected ABI statements and their associated final exam questions is 63%, which still leaves 37% of the responses inconsistent between the two instruments. For comparison, the overall fraction of consistent responses between the selected ABI statements and their associated *prelim* questions is 68%. Between the ABI and either the prelims or the final exam, the consistency in either case is about two-thirds, which supports my use of the BHGP to check for consistent responses.

The tendency for students to respond consistently also depends on the ABI statement-final exam question pair. For example, student responses to sA78n, “the

Moon causes part of the tides," were consistent with their responses to the associated exam question 90% of the time. At the lowest end of the spectrum, student responses to sA85n, "the Moon rotates even though we see only one side of it," were consistent with their responses to the associated exam question only 35% of the time. A total of 16 of the 20 ABI statement-final exam question pairs had response consistencies ranging from 52% to 73%.

4.4.2 Final exam and ABI response consistency vs. childhood interest in astronomy

Using the flow chart in Figure 4.11 and the 20 comparisons between final exam questions and associated ABI statements, presented in Appendix G, for each student, I determined the fraction of consistent responses. I then compared the fraction of consistent responses with their childhood interest in astronomy. I found that the correlation between the fraction of consistent responses and childhood interest in astronomy is 0.357 ($r^2 = 0.13$, $df = 55$, $p = 0.006$), which represents that 13% of the variance in consistent misconception recollection is accounted for by the variance in childhood interest in astronomy. The result is statistically significant, indicating that *the likelihood that ABI scores can predict student performance on the final exam may increase with childhood interest in astronomy*. In other words, for those students who had a relatively low childhood interest in astronomy, their scores on the ABI poorly predict their performance on the final exam. On the other hand, for those students who had a relatively high childhood interest in astronomy, their scores on the ABI are more likely to make a reliable prediction of their performance on the final exam. Figure 4.12 presents a plot of the fraction of consistent responses between the ABI and the final exam vs. childhood interest in astronomy.

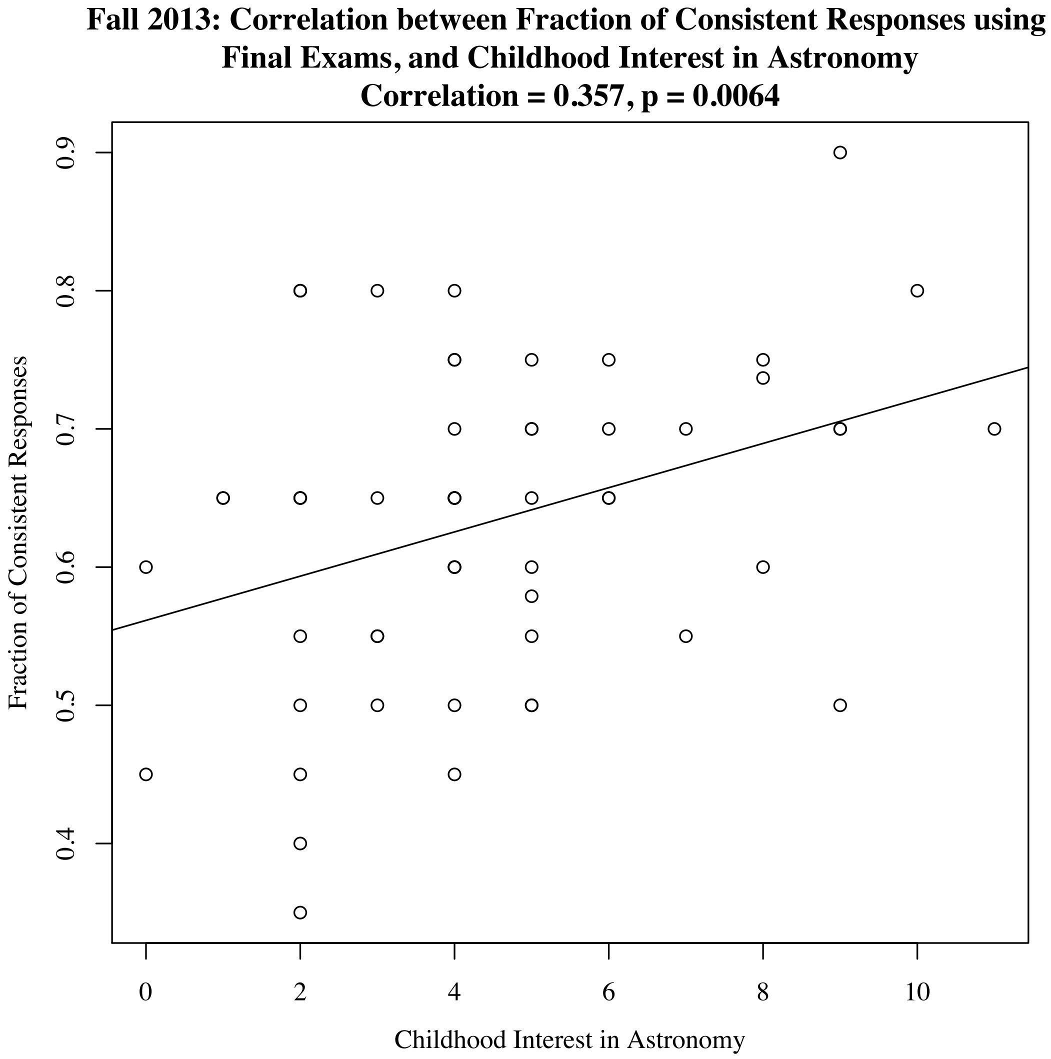


Figure 4.12. Fraction of consistent responses between the ABI and the final exam vs. childhood interest in astronomy

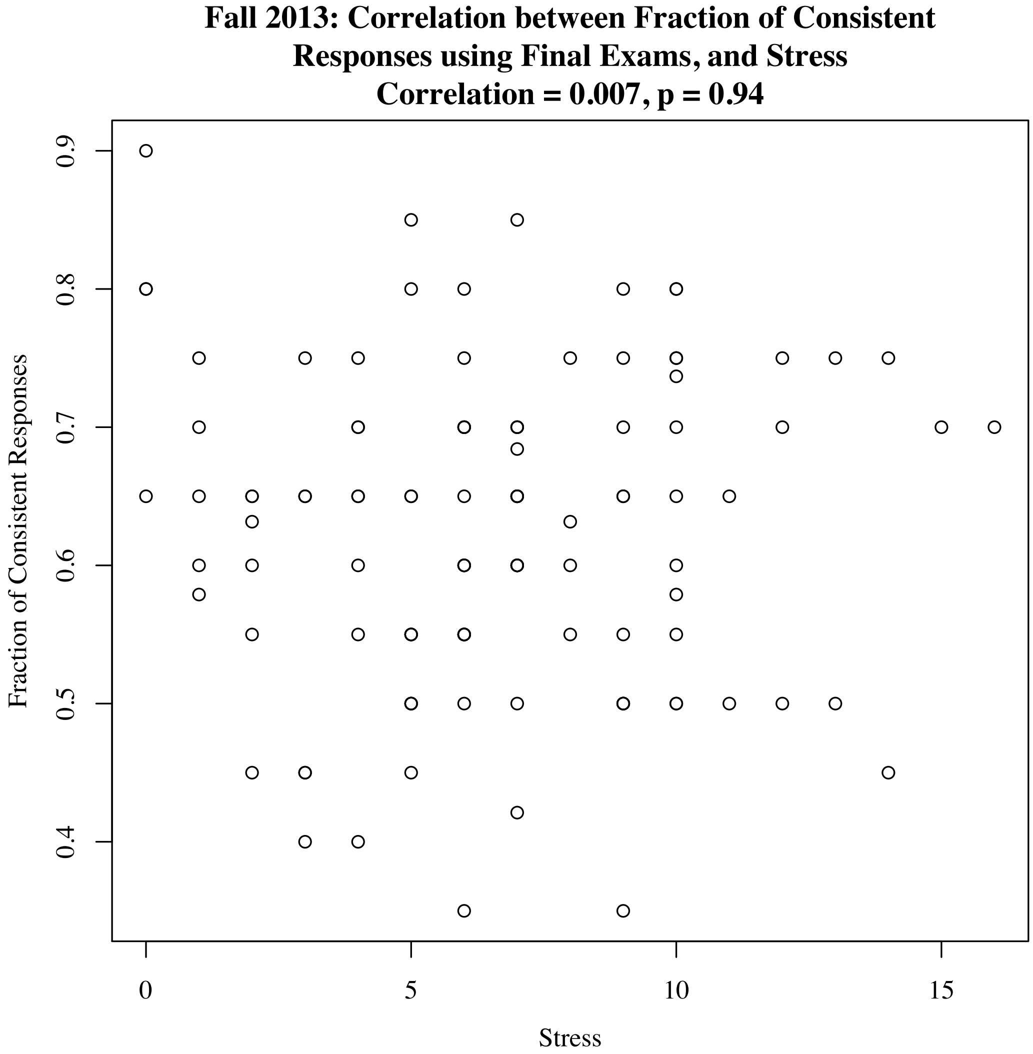


Figure 4.13. Fraction of consistent responses between the ABI and the final exam vs. stress

4.4.3 Final exam and ABI response consistency vs. stress

I then compared the fraction of consistent responses with the self-reported stress level of the students. I found that the correlation between the fraction of consistent responses and stress is −0.007 ($r^2 < 0.01$, $df = 105$, $p = 0.94$), which represents that less than 1% of the variance in consistent misconception recollection is accounted for by the variance in stress. The result, which is not at all statistically significant, indicates that stress, as measured, is unrelated to response consistency between the ABI and the final exam. This result is consistent with those in Sections 4.2.3 and 4.3.3. Figure 4.13 presents a plot of the fraction of consistent responses between the ABI and the final exam vs. childhood interest in astronomy.

4.5 From endorsement of the correct science to denying it

In Appendix F, I present questions that are associated with particular statements in the ABI. Likewise, in Appendix G, I present final exam questions that are associated with particular statements in the ABI. Some questions between the two sets overlap, in that they probe the same misconception and have nearly identical wording and response options. Hence, I can compare student responses to the relevant questions from both instruments to estimate the fraction of the time that students go from endorsing the correct science, at the time of a prelim, to denying the correct science on the final exam. Outside of guessing on exam questions, of which there is a 20% chance per question under consideration, the fraction of the time that students go "from correct to incorrect" *approximates the rate of recollection inconsistency*, because students who endorse the correct science in the course generally ought to remember it by the end of the course. Appendix H presents the overlapping prelim and final exam questions. In total, I made eight comparisons. On average, I find that *those students who endorse the correct science on a prelim denied it 26% of the time on the final exam*.

4.6 Attendance questions

4.6.1 Description

As discussed in Section 2.1.1, at the end of each class, NFC asks a misconception-based attendance question about a topic to be lectured in the subsequent class day preceding the topic. A selected set of these "attendance questions" serves as yet another pre-instruction probe for checking the consistency of student recollections of their past misconceptions. For the subsequent analysis, I am not interested in whether or not a student necessarily answers an associated *exam* question correctly. Instead, I can check to see whether or not students' written responses may be identified with endorsement of the correct science. For clarification, responses to attendance questions may be brief phrases, rather than explanations, since I am not interested in quantifying the depth to which students can explain their reasoning. Rather, I am interested in assessing the consistency of one's recollection of their own past beliefs.

For each attendance question, I determined categories that best capture the correctness of the responses, as described in Table 4.5. For each attendance question, I worked with NFC to establish specific categories. In total, I made 12 comparisons between attendance questions and their associated ABI statements. Appendix I presents the attendance questions under consideration, their specific categories, their associated ABI statements, the number of student responses to each attendance question, and the fraction of student responses to each category of answers. Included with the set of attendance questions in Appendix I is Fleiss' κ for each question. A brief discussion of the meaning of Fleiss' κ follows shortly.

Note that for attendance questions 4, 8, and 9, I divided typical correct responses into two further categories: "correct," and "nearly correct." I consider nearly correct responses as those which are closer to the correct answer than to a common misconception of interest. For the purpose of comparing attendance question responses

Category	Representation
1	a typical correct response by a student
2	a typical incorrect response that is associated with one or more *common* misconceptions, such as those already present in the ABI
3	all other “incorrect” responses

Table 4.5. Attendance question categories

with those of the associated ABI statements, I considered a response “correct” if it was at least “nearly correct,” and “incorrect” otherwise.

The challenge to scoring the written responses to the attendance questions, however, is in appropriately categorizing the responses. With appropriate categories established, a standard procedure for checking the consistency of the categories is to perform an analysis of inter-rater reliability (Eldep Jr., Pavalko, & Clipp, 1993, pp. 42–44). For a particular attendance question, I score a sample of written responses. Then I present the same sample to an independent rater, that is, a second person who scores the responses independently. The independent rater is provided the same *categories*, but does not see my scores. The agreement between the two raters, represented by Fleiss' κ, is given by

$$\kappa = \frac{\text{overall agreement} - \text{chance agreement}}{1 - \text{ chance agreement}}, \tag{4.1}$$

where overall agreement is the fraction of scores which agree between the two raters. The chance agreement reflects the tendency for two inter-raters to have the same total number of responses in each category, even though the actual responses themselves may be in different categories between the two raters. Mathematically, for a sample of N responses and C categories, chance agreement is given by

$$\text{chance agreement} = \frac{1}{N^2} \sum_{c=1}^{C} R1_c \, R2_c, \tag{4.2}$$

where $R1_c$ is the number of responses that rater #1 scored as belonging to category c, and $R2_c$ is the number of responses that rater #2 scored as belonging to the same category. Values of $\kappa > 0$ indicate that agreement is better than chance, with $\kappa = 1$ representing perfect agreement. Table 4.6 presents guidelines for the interpretation of values of κ; according to Landis and Koch (1977), $\kappa > 0.4$ represent at least moderate agreement.

κ	**Interpretation**
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Table 4.6. Interpretation of Fleiss' kappa

With the aid of inter-rater Jessica W. Clark, graduate Ph.D. student in physics education research at the University of Maine, I conducted an inter-rater analysis for a subset of four of the nine attendance questions under consideration. For each analysis, I randomly selected a subset of 25 responses. Values of κ ranged from 0.62 to 1.00, which, according to Table 4.6, indicates at least substantial agreement between the inter-raters. After scoring the responses, we underwent a brief period of discussion about the results and suggestions for category revisions. The final categories for the attendance questions are those presented in Appendix I.

Having established that the codes that I assigned to the attendance question responses agree reasonably with those of an inter-rater, I may continue with my analysis of recollection reliability. Because each attendance question essentially serves the same purpose as a pretest question, as with any of the BHGP questions, I can use the attendance questions and the associated ABI statements to determine the appropriate category of memory recollection, as described in Section 4.2.1. The flow chart with which I determine the consistency of one's recollections otherwise proceeds in the same manner as illustrated in Figure 4.3.

Respectively, the overall percentages of definitely consistent recollections and definitely *in*consistent recollections are 30% and 18%. These percentages are nearly identical to those associated with recollections of black hole and galaxy misconceptions, at 34% and 20%, respectively (refer to Section 4.2.1), which shows that the scoring of attendance question responses produces a similar calculated fraction of recollection inconsistency as does a multiple-choice-designed pretest.

The tendency for students to respond consistently also depends on the attendance question-ABI statement pair. For example, student responses to sA189, “the Sun is the brightest star in universe,” were definitely consistent with their responses to the associated attendance question 56% of the time. At the lowest end of the spectrum, I was unable to measure any definite consistency between student responses to sA248, “cosmic rays are light rays,” and their responses to the associated attendance question, because only three students in the entire class provided a correct description of cosmic rays prior to instruction. A total of 9 of the 12 attendance question-ABI statement pairs had *definite* response consistencies ranging from 20% to 47%.

4.6.2 Attendance question recollection consistency vs. childhood interest in astronomy

Using the 12 comparisons between attendance questions and associated ABI statements, presented in Appendix I, for each student, I determined the fraction of both consistent and inconsistent misconception recollections, as well as the fraction of the time in which I cannot tell either way. I then compared the fraction of both consistent and inconsistent attendance question misconception recollections with their childhood interest in astronomy.

First, I compared the fraction of consistent recollections with their childhood interest in astronomy. I found that the correlation between the fraction of consistent

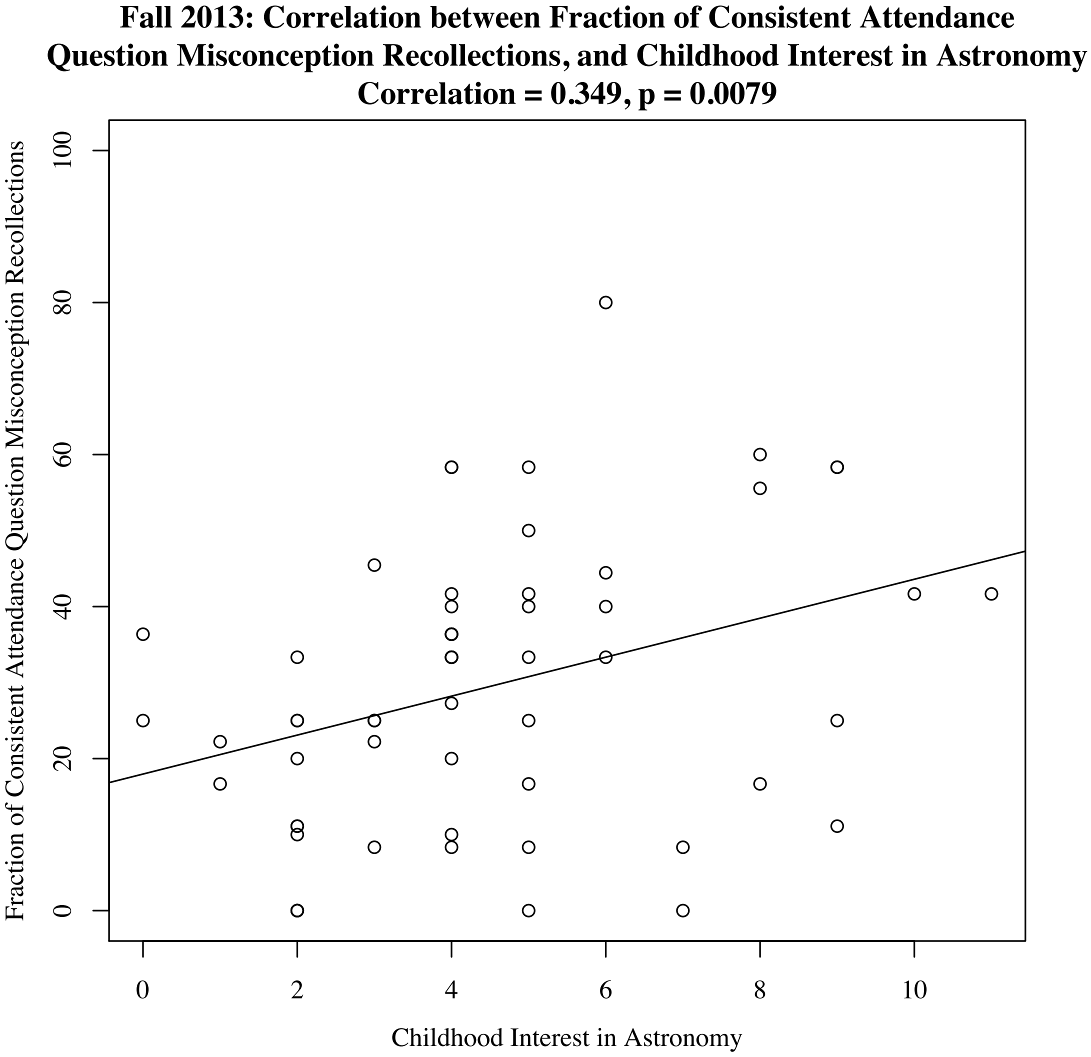


Figure 4.14. Fraction of consistent attendance question misconception recollections vs. childhood interest in astronomy

recollections and childhood interest in astronomy is 0.349 ($r^2 = 0.12$, $df = 55$, $p = 0.008$), which represents that 12% of the variance in consistent misconception recollection is accounted for by the variance in childhood interest in astronomy. The result is statistically significant, indicating tentatively that increased childhood interest in astronomy is correlated with increased accurate recollection of astronomy misconceptions. Figure 4.14 presents a plot of the fraction of consistent attendance question misconception recollections vs. childhood interest in astronomy.

Then I compared the fraction of *in*consistent recollections with their childhood interest in astronomy. I found that the correlation between the fraction of *in*consistent recollections and childhood interest in astronomy is −0.045 ($r^2 < 0.01$, $df = 55$, $p = 0.74$), which represents that less than 1% of the variance in *in*consistent misconception recollection is accounted for by the variance in childhood interest in astronomy. The result is not statistically significant, indicating that increased childhood interest in astronomy does not cause any significant increase in *in*accurate recollection of astronomy misconceptions. Figure 4.15 presents a plot of the fraction of consistent attendance question misconception recollections vs. childhood interest in astronomy. Together, these results suggest that *there is a significant correlation between childhood interest in astronomy and definitely consistent memory recollections*, a result which clarifies that of Section 4.2.2.

4.6.3 Attendance question recollection consistency vs. stress

I then compared the fraction of both consistent and inconsistent attendance question recollections with the self-reported stress level of the students. First, I compared the fraction of consistent recollections with their stress. I found that the correlation between the fraction of consistent recollections and stress is 0.014 ($r^2 < 0.01$, $df = 107$, $p = 0.89$), which represents that less than 1% of the variance in consistent misconception recollection is accounted for by the variance in stress. The

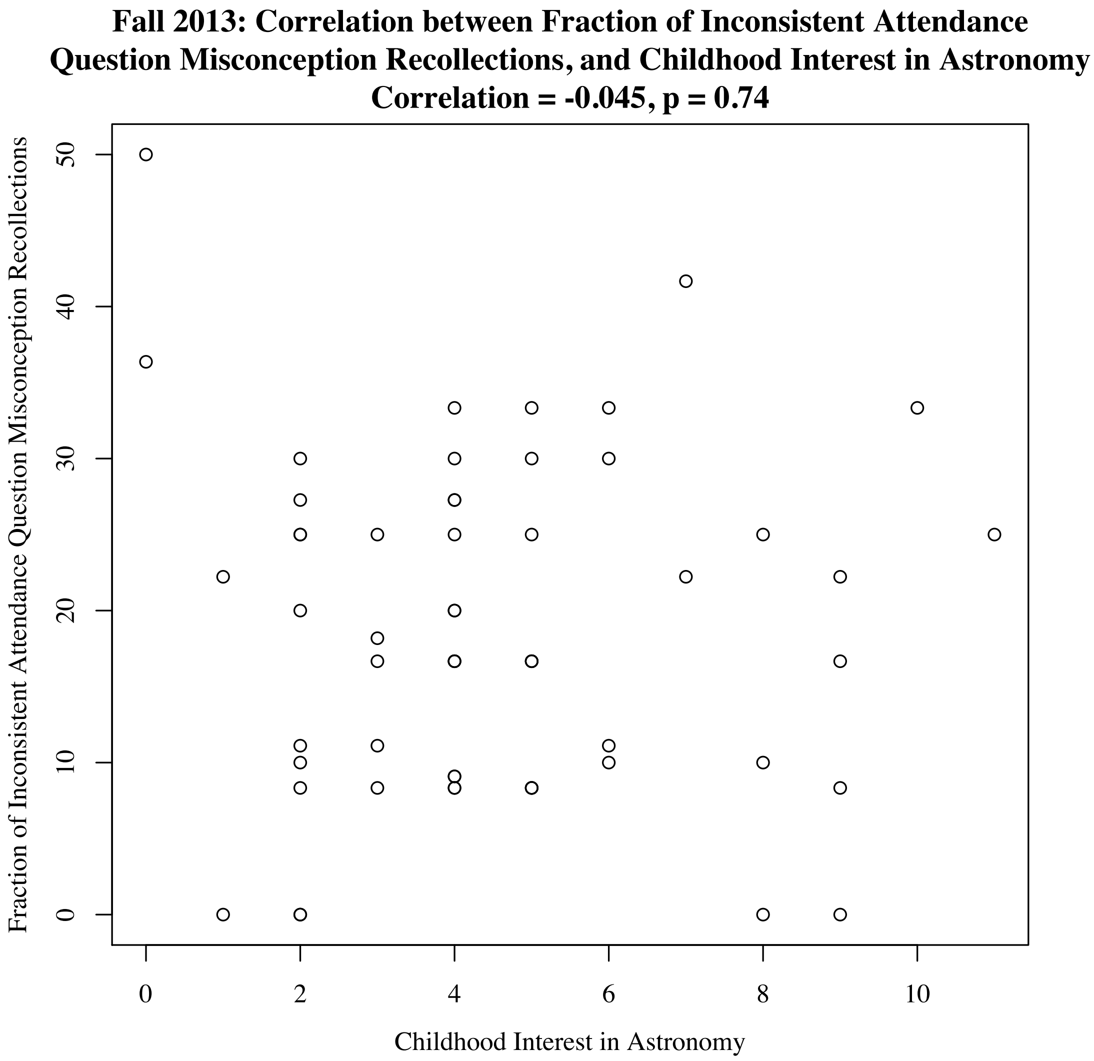


Figure 4.15. Fraction of inconsistent attendance question misconception recollections vs. childhood interest in astronomy

result is not at all statistically significant, indicating tentatively that stress level does not seem to affect recollection consistency. Figure 4.16 presents a plot of the fraction of consistent attendance question misconception recollections vs. stress.

Then I compared the fraction of *in*consistent recollections with the self-reported stress level of the students. I found that the correlation between the fraction of *in*consistent recollections and stress is -0.075 ($r^2 < 0.01$, $df = 107$, $p = 0.44$), which represents that less than 1% of the variance in *in*consistent misconception recollection is accounted for by the variance in stress. The correlation is negative, which would suggest a relationship between *decreased* stress and inaccurate recollection. The result, however, is not statistically significant, indicating that decreased stress does not cause any significant increase in *in*accurate recollection of astronomy misconceptions. Figure 4.17 presents a plot of the fraction of consistent attendance question misconception recollections vs. childhood interest in astronomy. Together, these results suggest that *stress does not significantly influence memory recollection consistency*. This result is also consistent with that of Section 4.2.3, in which I found no significant influence of stress on one's own recollection of black hole and galaxy misconceptions.

4.7 Discussion of results

In this Chapter, I have introduced surveys related to childhood interest in astronomy and stress as two independent probes to measure the extent to which childhood interest in astronomy and stress influence the consistency of memory recollection. I have also introduced a pretest on black holes and galaxies to assess the extent to which students endorse misconceptions related to black holes and galaxies prior to instruction. These instruments, in conjunction with prelims, the final exam, and the ABI, can be used to assess the validity of the ABI as a tool for examining misconceptions in the course.

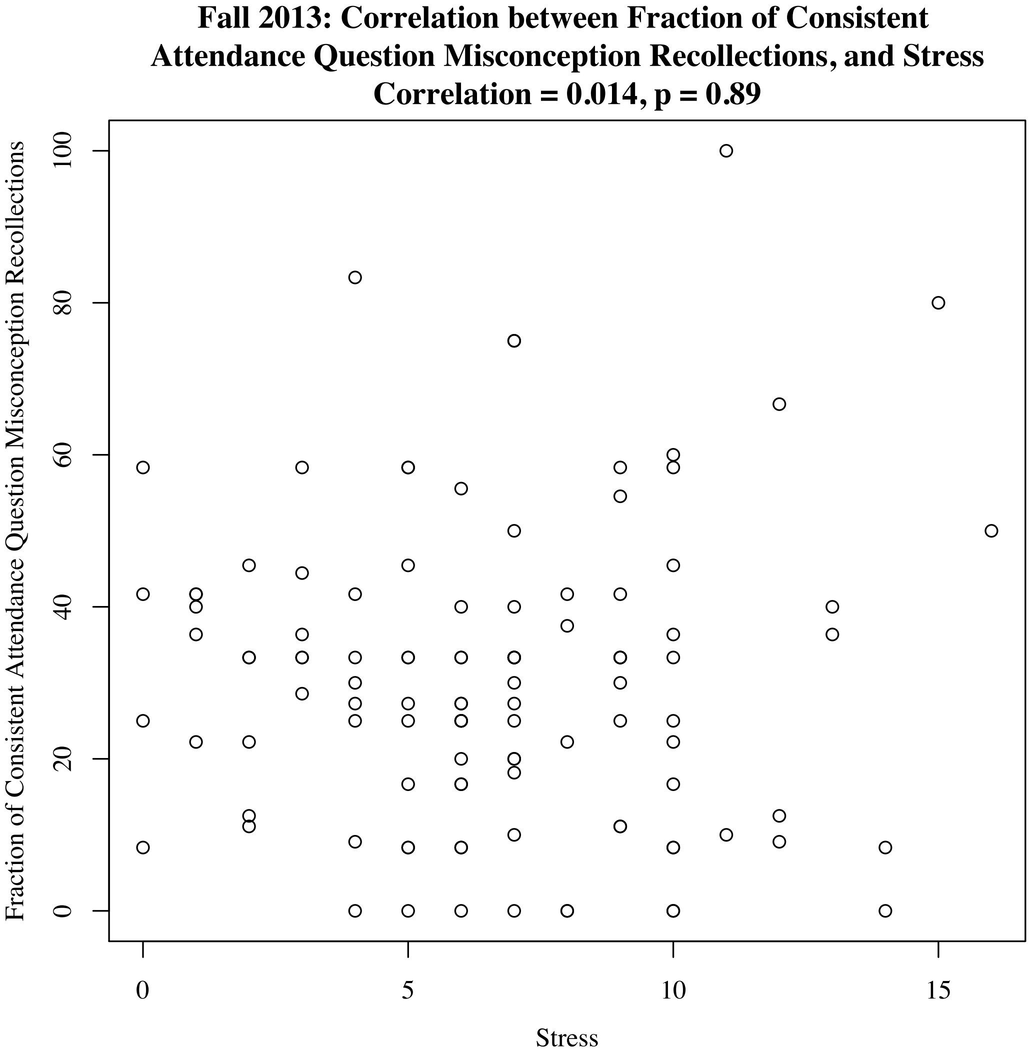


Figure 4.16. Fraction of consistent attendance question misconception recollections vs. stress

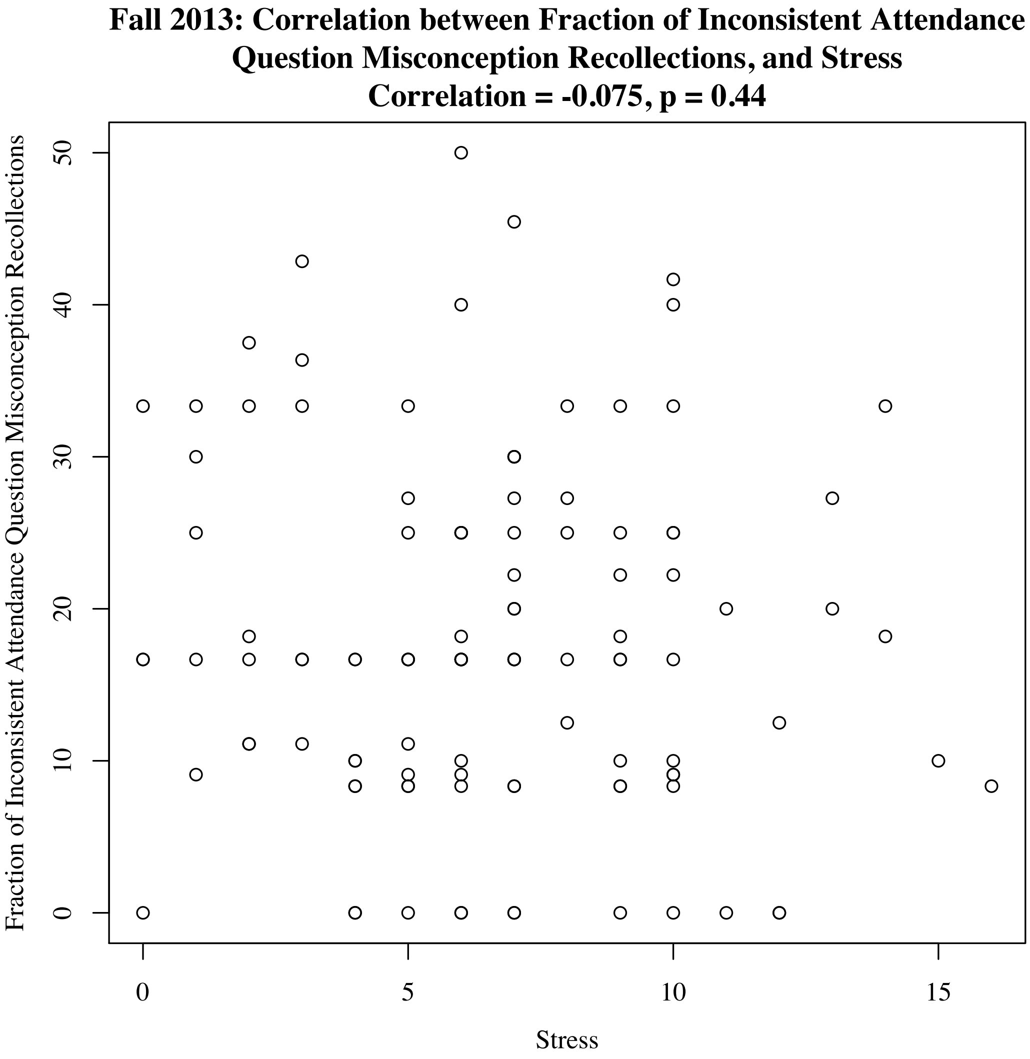


Figure 4.17. Fraction of inconsistent attendance question misconception recollections vs. stress

With regard to childhood interest in astronomy and stress, the results for this Chapter are as follows:

1. One can measure childhood interest in astronomy independently from one's stress level (Section 4.1.3).
2. There is an overall marginal correlation between childhood interest in astronomy and consistent memory recollections (Sections 4.3.2 and 4.6.2).
3. The likelihood that ABI scores can predict student performance on the final exam increases with childhood interest in astronomy (Section 4.4.2).
4. Stress has no significant influence on recollection consistency (Sections 4.2.3, 4.3.3, 4.4.3, and 4.6.3).

With regard to the consistency of recollections, as exemplified by comparing the data between multiple instruments, the results for this chapter are as follows:

1. Recollections of black hole and galaxy misconceptions are *definitely* inconsistent about 20% of the time (Section 4.2.1).
2. Responses between prelim questions and their associated ABI items are inconsistent 32% of the time (Section 4.3.1).
3. Responses between final exam questions and their associated ABI items are inconsistent 37% of the time (Section 4.4.1).
4. Students who endorse the correct science on a prelim deny it on the final exam 26% of the time (Section 4.5).
5. Responses between pre-instruction attendance questions and their associated ABI items are *definitely* inconsistent 18% of the time (Section 4.6.1).

Note that both the BHGP and the attendance questions are instruments whose questions are deliberately designed to examine whether or not students endorse *misconceptions*, rather than the correct science, as on the prelims and the final exam. The percent of definitely *in*consistent recollections, as evidenced by responses to the BHGP and the attendance questions, ranges from 18% to 20%. For comparison, the probability of randomly guessing an answer to any question on the associated BHGP is 20%, which suggests that students are guessing on the ABI $\gtrsim$ 20% of the time.

Of particular note is the statistic that students who endorse the correct science on a prelim deny it on the final exam 26% of the time. For comparison, the percent of inconsistent recollections, as evidenced by responses to prelim and exam questions vs. those on the ABI, ranges from 32% to 37%. These instruments are based on the use of standard examination questions, rather than questions designed to probe misconceptions, as with the BHGP and the attendance questions. In general, the highest proportion of inconsistent responses is observed using data from prelims and the final exam, most notably considering that students deny the correct science that they endorsed on a prelim 26% of the time. Hence, the validity of the ABI as an instrument at assessing student misconceptions is *comparable* to that of an instrument which uses standard examination questions.

CHAPTER 5

DESCRIPTIVE STATISTICS ON THE ASTRONOMY BELIEFS INVENTORY DATA

In Chapter 3, I showed that the ABI is a valid instrument from which one can analyze incorrect beliefs held by college students in astronomy, past or present. In this Chapter, I present descriptive statistics on the ABI data. As a reminder to the reader, the data codes are described in Table 2.4 (page 28), such that a score of “1” represents the lowest degree of misconception retainment, and a score of “3” represents the highest degree of misconception retainment.

5.1 Overall inventory scores per semester

To examine the overall inventory scores, I first calculated a representative mean misconception retainment score for each student, as described in Section 3.1. For each semester, I then summed each of the mean misconception retainment scores per student, then divided by the semester sample size. To determine the spread of mean misconception retainment scores per student, I also calculated, per semester, the standard deviation of the mean student misconception scores. Table 5.1 reports the MMRS (refer to Section 3.1) and the standard deviation about the semester means.

Semester	Class Size	Sample Size	MMRS
Fall 2009	188	114	1.94 ± 0.24
Fall 2010	175	107	1.89 ± 0.22
Fall 2011	171	91	1.84 ± 0.28
Fall 2012	170	91	1.71 ± 0.24
Spring 2013	192	126	1.91 ± 0.25
Fall 2013	174	110	1.91 ± 0.25

Table 5.1. Mean misconception retainment scores per semester

The means reported in Table 5.1 range from 1.84 to 1.94 for five of the six semesters. For comparison, a score of 2 means that overall, students typically endorse misconceptions through adolescence, but not typically after instruction in the course. Hence, a range of scores from 1.84 to 1.94 mean that on average, students in these five semesters are more likely to abandon misconceptions than retain them even after instruction. In the Fall 2012 semester, the MMRS is 1.70, which indicates that those students who volunteered themselves took AST 109 having already dispelled themselves of a higher fraction of misconceptions than students in the other semesters.

To quantify the extent to which these scores are significantly different, I performed an ANOVA test on the MMRS for each of the six semesters. As discussed in Section 3.3, an ANOVA test compares the variance between groups (in this case, semesters) vs. within each group. A statistically significant result indicates that the mean of the scores in at least one semester differs significantly from the mean of the scores in other semesters. With regard to the homogeneity of variances, the Levene statistic $W = 1.586$, and the significance of the differences in the variances $p_V = 0.162$, indicating that the variances of the semester misconception retainment were not significantly different. The actual semester means, however, were significantly different ($F(5,633) = 10.97$, $p < 0.0005$, $\eta^2 = 0.080$, $\omega^2 = 0.072$), indicating that overall student performance on the ABI varied significantly from semester to semester.

One might expect, at least, that the MMRS for the first four semesters (Fall 2009 to Fall 2012) ought to be about the same, for a number of reasons. First, the teaching pedagogy of NFC remained consistent during this time frame. Second, the format of the ABI remained consistent. And, third, the ABI was administered consistently at the end of the semester, for each of the semesters. To test that the MMRS for the first four semesters were the same, I performed an ANOVA test on the MMRS for the Fall 2009 to Fall 2012 semesters. With regard to the homogeneity of variances, the Levene statistic $W = 2.66$, and the significance of the differences in the variances $p_V = 0.048$, indicating that the variances of the semester misconception retainment means are marginally significantly different. The actual semester means, however, were significantly different ($F(3, 399) = 15.50$, $p < 0.0005$, $\eta^2 = 0.104$, $\eta^2 = 0.097$), indicating that despite the consistency of the ABI administration and teaching pedagogy, overall student performance on the ABI varied significantly.

The significantly differing scores among the first four semesters could be attributed to a variety of reasons. For example, in the first four semesters, only one time slot was made available to the students to complete the ABI. The time slot was optimally chosen by NFC to permit the largest proportion of the class to complete the ABI voluntarily. Students who were unable to make the time slot had to make arrangements to meet with NFC. Overall, 59% of the students from Fall 2009 to Fall 2012 volunteered themselves to respond to the ABI, and, 64% of the students from the Spring 2013 and Fall 2013 semesters volunteered. The MMRS for the Fall 2012 semester is also lower than the MMRS in the preceding three semesters, which may be attributed to selection effects, wintry travel conditions for the night of administration, or other conditions.

To calculate the fraction of misconceptions endorsed even after instruction in the course, I recoded the data into two categories: one for endorsing the incorrect belief even after instruction, and one for unlearning the incorrect belief anytime before the end of the course, as indicated in Table 3.5. Interestingly, an investigation of the final exam scores of the student volunteers, from Fall 2009 to Fall 2012, reveals no trend in final exam scores from semester to semester, even though the final exams themselves remained virtually identical to each other. The mean final exam scores of the participants, for each semester, are, respectively, 63, 62, 64, and 63. Missing cases were excluded listwise, which is to say for a particular student if a student did not take either the exam or the ABI, then all data for that student were ignored in the analysis. The mean fractions of misconceptions retained by those students from Fall 2009 to Fall 2012 are, respectively, 0.20, 0.17, 0.19, and 0.12. Hence, the selection effect due to the voluntary nature of the study may not be so serious as to preclude the use of the acquired data in selected further statistical analyses. For example, as I show in Chapter 7, my analysis of *correlations* between statements in the ABI produces reliable results provided that I incorporate data from all six semesters (inc. Spring 2013 and Fall 2013), rather than just use data from exclusively a particular semester.

To illustrate the relationship between misconceptions endorsed in the course and performance on the final exam, I correlated the fraction of misconceptions endorsed by the students, in the Fall 2009 to Fall 2012 semesters, with their final exam scores. The correlation between fraction of misconceptions endorsed and final exam score has the value -0.525 ($r^2 = 0.28$, $df = 387$, $p < 0.0005$). Therefore, *the fraction of misconceptions that students report having dispelled as a result of instruction in the course can be used as a generalized predictor of their approximate performance on the final exam*. Of the first four semesters under present consideration, there is a selection effect due to students volunteering themselves to respond to the ABI. Therefore, I cannot assess the extent to which this correlation holds for the portion of the students who did not volunteer themselves (about 41% of the overall study).

5.2 Analysis of response code frequencies per semester

To investigate further the mean misconception retainments of students, I analyzed the frequencies of the response codes (1, 2, 3) for all students in each of the six semesters. As explained in Section 2.2, some students in the Fall 2012 semester responded to only 187 statements. To ensure an unbiased analysis of the frequencies in the subsequent discussion, these students have been omitted. In the event that a student responded to fewer than 215 items, I normalized the frequency of each response code so that their sum added to 215. Table 5.2 presents the mean number of 1s, 2s, and 3s per semester, along with the standard deviation (SD) about the means. As a reminder, in the Spring 2013 semester, taught by DJB, and the Fall 2013 semester, taught by NFC, the inventory format was changed (see Table 2.3).

With regard to the data from the Fall 2009 to Fall 2012 semesters, Table 5.2 shows that, as suspected, students reported having unlearned a higher number of misconceptions prior to college (response code "1") in the Fall 2012 semester (87) compared to students in the earlier semesters (in reverse chronological order: 74, 61, 55). While the variances among the reported number of misconceptions unlearned prior to college are significantly different from each other ($W = 4.067$, $p_V = 0.007$), an ANOVA test nonetheless confirms that the mean number of misconceptions unlearned prior to college differs significantly among the semesters ($F(3,395) = 12.63$, $p < 0.0005$, $\eta^2 = 0.084$, $\omega^2 = 0.080$).

With regard to the data from the Fall 2009 to Fall 2012 semesters, Table 5.2 shows that the number of misconceptions that students retain, even after instruction (response code "3"), is the lowest for the Fall 2012 semester (24) and higher for the earlier semesters (in reverse chronological order: 40, 37, 43). The variances among the semester scores are marginally significantly different ($W = 2.185$, $p_V = 0.089$). An ANOVA test confirms that the mean number of misconceptions retained even after instruction in AST 109 differs significantly among the semesters ($F(3,395) = 13.51$, $p < 0.0005$, $\eta^2 = 0.093$, $\omega^2 = 0.086$). For comparison, Table 5.2 shows that the mean number of misconceptions dispelled in AST 109 (response code "2") also differs significantly among the semesters ($F(3,395) = 4.85$, $p = 0.003$, $\eta^2 = 0.036$, $\omega^2 = 0.028$) though to a lesser extent, which indicates that there is some significant variability in the scoring throughout the first four semesters.

In contrast, Table 5.2 paints a different picture for student performance in the Spring 2013 and Fall 2013 semesters. Table 5.2 indicates that the mean number of misconceptions dispelled in AST 109 (response code "2"), for Spring 2013 and Fall 2013, is the lowest (respectively, 79 and 88) compared to those students taught under

Semester	**Sample Size**	**Freq. # 1s**	***SD***	**Freq. # 2s**	***SD***	**Freq. # 3s**	***SD***
Fall 2009	114	55.5	36.6	116.3	33.1	43.2	23.3
Fall 2010	107	61.4	33.3	116.5	33.2	37.1	23.4
Fall 2011	91	73.9	44.9	101.2	39.1	40.0	23.5
Fall 2012	87	87.2	41.9	104.2	39.9	23.6	19.5
Spring 2013	126	77.4	37.6	78.6	32.4	59.1	22.6
Fall 2013	110	73.0	40.5	87.8	36.2	54.1	21.6

Table 5.2. Mean frequency and standard deviation (*SD*) of each response code to the ABI

NFC from Fall 2009 to Fall 2012 (104-116). As further confirmation, Table 5.2 indicates that the mean number of misconceptions endorsed, even after instruction, in AST 109, for Spring 2013 and Fall 2013, is the highest (respectively, 59 and 54) compared to those students taught under NFC from Fall 2009 to Fall 2012 (24-43). Recall that the ABI format was changed for the Spring 2013 and Fall 2013 semesters, but within those two semesters, the same instrument was employed. These results strongly suggest that changes to the inventory format more greatly affect student reports of their own misconceptions than differences in teaching pedagogy. In other words, the way that students respond significantly depends on the context of the list of statements. This result is consistent with that from my analysis in Section 3.8 and again later in Section 7.2.4.

5.3 Analysis of misconception retainment by gender

As a further investigation of the mean semester misconception retainment scores, I examined the differences in scores based on gender for students under NFC (Fall 2009 to Fall 2012, and Fall 2013) and students under DJB (Spring 2013). I performed separate one-way ANOVA tests, one test per instructor, on the mean semester misconception retainment scores for both genders. The semester groups under consideration are Fall 2009 to Fall 2012, Spring 2013, and Fall 2013. Among the groups, respectively, there are 202, 72, and 67 male subjects and 200, 54, and 43 female subjects. The variances of the scores between the genders are significantly different for only the Fall 2013 semester (respectively, $p_V = 0.26$, $p_V = 0.98$, and $p_V = 0.01$). Table 5.3 reports the MMRS, as defined in Section 3.1, separately for both male and female students, where values of p ranging from 0.0005 to 0.05 are marked with an asterisk (*), and values of $p < 0.0005$ are marked with a double asterisk (**).

Table 5.3 shows that for both instructors, female students are significantly more likely to retain misconceptions in astronomy than male students, because the MMRS

Semester(s)	**Gender**	***n***	**MMRS**	**Std. Dev**	F	p	η^2	ω^2
Fall 2009–Fall 2012:	**Male**	202	1.79	0.26	29.31	<0.0005**	0.068	0.066
	Female	200	1.92	0.24				
Spring 2013:	**Male**	72	1.86	0.23	10.44	0.002*	0.078	0.070
	Female	54	1.99	0.24				
Fall 2013:	**Male**	67	1.87	0.28	4.26	0.041*	0.038	0.029
	Female	43	1.97	0.19				

Table 5.3. Mean semester misconception retainment, by gender

for females is significantly higher than for males. To investigate whether or not female students tend to retain more misconceptions than male students, I recoded all of the ABI data into two categories: one for retaining the incorrect belief even after instruction, and one for unlearning the incorrect belief anytime before the end of the course, as indicated in Table 3.5. Using the recoded data, I calculated the mean fraction of misconceptions that male and female students endorsed under both instructors. I then performed a one-way ANOVA test on the recoded scores. The variances of the scores between the genders are significantly different for only the Fall 2009 to Fall 2012 semester group (respectively, $p_V = 0.04$, $p_V = 0.28$, and $p_V = 0.62$). The results are presented in Table 5.4, where "Misc. Endorsed" is the mean fraction of misconceptions endorsed. The data suggest that even after instruction, female students may be more likely to hold on to misconceptions in astronomy than male students.

Is it possible that astronomy misconceptions tend to persist more with female students than male students specifically because female students learned less of the correct science as children and adolescents? To answer this question, I recoded the ABI data into two categories: one for endorsing the incorrect belief through adolescence, and one for disabusing oneself of the incorrect belief during childhood or adolescence, as indicated in Table 4.2. I then calculated the mean fraction of misconceptions that male and female students, under both instructors, endorsed through their respective adolescences. I then performed a one-way ANOVA test on the recoded scores. Interestingly, the variances of the scores between the genders were significantly different from each other within two of the three sets (respectively, $p_V = 0.01$, $p_V = 0.63$, and $p_V < 0.01$). The results are presented in Table 5.5, where "Misc. Through Adol." is the mean fraction of misconceptions endorsed through adolescence.

Despite the non-homogeneity in variances between genders, the data in Table 5.5 indeed support lower misconception endorsement through adolescence for males

Semester(s)	Gender	n	Misc. Endorsed	Std. Dev	F	p	η^2	ω^2
Fall 2009–Fall 2012:	**Male**	202	0.139	0.100	37.17	<0.0005**	0.085	0.083
	Female	200	0.203	0.111				
Spring 2013:	**Male**	72	0.259	0.108	3.89	0.051	0.030	0.022
	Female	54	0.296	0.098				
Fall 2013:	**Male**	67	0.248	0.105	0.31	0.581	0.003	–0.001
	Female	43	0.258	0.094				

Table 5.4. Mean fraction of misconceptions endorsed even after instruction, by gender

Semester(s)	**Gender**	***n***	**Misc. Through Adol.**	**Std. Dev**	F	p	η^2	ω^2
Fall 2009–Fall 2012:	**Male**	202	0.647	0.205	14.36	<0.0005**	0.035	0.032
	Female	200	0.718	0.165				
Spring 2013:	**Male**	72	0.597	0.165	11.16	0.001	0.083	0.075
	Female	54	0.698	0.172				
Fall 2013:	**Male**	67	0.626	0.210	6.07	0.015	0.053	0.044
	Female	43	0.714	0.134				

Table 5.5. Mean fraction of misconceptions endorsed through adolescence, by gender

than for females. In other words, male students either learn more of the correct science or tend to disabuse themselves of misconceptions in astronomy more easily during their youth than female students. If male students learn more of the correct science or tend to let go of misconceptions in astronomy more easily during their youth than female students, then one should expect that female students will have held on to a significantly higher fraction of misconceptions through adolescence. The results in Table 5.5 are consistent with this hypothesis, which suggests that female students tend to retain more misconceptions through AST 109 because of the lack of instruction during their childhood and adolescence.

Could the discrepancy in mean fraction of misconceptions endorsed between male and female students be explained away by a discrepant degree of childhood interest in astronomy between the genders? To answer this question, I performed an ANOVA test on the scores to the CIAS (refer to Section 4.1.1) and grouped the students from the Fall 2013 semester together by gender. The variances of the scores between the groups were not at all significantly different from each other ($W = 0.070$, $F(1,55) = 0.006$, $p = 0.94$), which indicates that both male and female students may have shared very similar childhood interests in astronomy. Hence, the discrepancy in mean fraction of misconceptions endorsed between male and female students cannot be explained away by a preference for students of a particular gender to have a higher childhood interest in astronomy.

In summary, my analyses are consistent with the hypothesis that *female students may retain more misconceptions in astronomy than male students at all stages in their lives*. As described in an unpublished Ph.D. dissertation, Gray (2006) investigated misconceptions in astronomy by boys and girls separately in 8th grade by way of a multiple choice test based on misconceptions in astronomy. Gray showed that there was a statistically significant difference in performance on the test between the genders, and that through interviews, both boys and girls had different levels of astronomy knowledge and different backgrounds. The results by Gray are thus consistent with the results from my analyses.

5.4 Summary of the overall inventory scores

Each administration of the inventory from Fall 2009 to Fall 2012 used the same inventory format. The inventory was administered under the same instructor who used the same teaching pedagogy each semester. Nevertheless, based on my analysis of the inventory responses, summarized in Table 5.2, one should expect that student reports to the inventory each semester will have a lot of variation in them in part due to selection effects. Since not all students in the class volunteered themselves, the data may not fairly represent the misconceptions dispelled or retained by the rest of the class. In addition, based on the data from those students who volunteered themselves, the wording of statements as true vs. false in the ABI (Section 3.5) has a more significant effect on student self-reports of misconception endorsement than does teaching pedagogy. I have also shown that female students tend to both bring more astronomy misconceptions to the college classroom and retain more misconceptions through the end of the course than male students.

CHAPTER 6

FACTOR ANALYSIS

In this Chapter, I examine correlations between misconceptions associated with statements in the ABI, to see which misconceptions students tend to unlearn together. The motivation for finding correlated misconceptions is for instructors to design astronomy lectures most in accordance with the way that students unlearn their misconceptions most effectively. I present an introduction to factor analysis methodologies and the role of principal components analysis in my analysis of misconceptions in the ABI. My discussion of factor analysis, in this Chapter, is largely taken from the presentation of factor analysis by Lee and Ashton (2007).

6.1 Overview of factor analysis

6.1.1 Construction of factors

In factor analysis, one is primarily concerned with reducing the total number of variables to a smaller set. A *factor* is an unobserved variable that represents one or more measured variables in a set. A measured variable that is well represented by the factor is said to have a high *loading* on the factor. A correlated subset of items, each on the same factor, with the highest loadings is considered a *group*. Factor analysis serves the purposes of (i) creating groups of statements within each topic (e.g., galaxy subtopics) and (ii) reducing the number of items in the inventory to several groups, each with fewer items than the total number of statements.

Factor analysis unveils groups of correlated misconceptions by calculating correlations within the data. For example, given a particular set of statements, taken from the ABI, if students who endorse a misconception associated with one of the statements within the set, then the students are also likely to endorse misconceptions associated with other statements in the set. Using the codes in Table 2.4, this is to say the following: given a set of statements that have correlated misconceptions associated with them, (i) students who respond with "1" to any one of the statements are also likely to respond with "1" to the others, (ii) students who respond with "2" to any one of the statements are also likely to respond with "2" to the others, and (iii) students who respond with "3" to any one of the statements are also likely to respond with "3" to the others. Comprehensively, misconceptions in a group are strongly correlated if students who respond to a statement with any of the three codes (1, 2, 3) are likely to respond with the *same* code to other statements in the group.

Some additional nomenclature is presented by Lee and Ashton (2007) to describe the fundamentals of factor analysis. Since a factor is considered a dimension that represents an unmeasured variable, each factor can represent a unique construct. For example, among the galaxy statements (of which there are 12, see Appendix A), the property "galaxy appearance" may represent one factor, or dimension, while "being in the center" may represent a different factor. While neither of these quantities were directly measured, they are inferred from the context of the groups that result from performing factor analysis. Since the subtopics represent a theme within the parent topic, each factor can be considered a theme that is best identified with the statements in the group.

In factor analysis, the set of factors is said to be *orthogonal* if no two factors are correlated with each other. Orthogonal factors can be visualized as a set of vectors drawn at right angles to each other. For $n > 3$ orthogonal factors, they are orthogonal to each other in an n-dimensional space. Orthogonal factors are especially useful for analyzing the ABI data, because the factor analysis presents groups of misconceptions that are best unlearned together, and each group represents a distinctly separate conceptual theme. The alternative to an orthogonal set is an *oblique* set, which occurs if at least one pair of factors is allowed to correlate. Oblique factors may be more useful for examining the clustering or the structure of the correlations themselves, rather than the grouping of subtopics within factors (Lee & Ashton, 2007). The format of the inventory has been modified as according to Table 2.3. For convenience, the statements in the ABI will be referred to as "items" henceforth. The "item" designation serves the same purpose of the "variable" designation in the statistical literature and will be used in the following overview of factor analysis, taken from Lee and Ashton.

In factor analysis, the correlation of responses between each pair of items a and b is calculated by the sample correlation coefficient C_{ab}, defined by Equation (3.4), on page 39. A correlation matrix is then constructed from the calculated correlations between each pair of items. Because factors are derived from those items that are the most strongly inter-correlated, the correlation matrix comprised of all the C_{ab} serves as the basis for performing factor analysis.

The construction of the first factor proceeds as follows. First, the item scores for all items in the set are standardized. Second, a linear combination of standardized scores for all the items is computed. The coefficients of the linear combination are determined based on the desired factor solution set (e.g., orthogonal vs. oblique). Third, the scores on the first factor are calculated from the aforementioned linear combination. Finally, the variance of the factor is calculated based on the correlations between the items and each factor. The correlations between items and factors are called *factor loadings*. The variance of the factor is called the *eigenvalue*. A factor with more strongly correlated items will have more variance and, thus, a larger eigenvalue.

With the first factor created, the variance of the first factor is removed from the original data set, so as to produce a second factor that is uncorrelated with the first. After the first factor is created, the second factor is determined in the same way, except with a different correlation matrix. The resulting correlation matrix, then, is turned into a matrix of *residual correlations*, with the residual being what remains after removing the variance of the preceding factor. The second factor is then created using otherwise the same process that was used to create the first factor. The third and subsequent factors are likewise created from the residual correlation matrix, whose residuals are calculated as a result of removing all variance associated with the preceding factors. This process continues until all the items in the analysis have been exhausted. The process of extracting uncorrelated factors from residual correlations is essentially the same process that is used in linear algebra to create an orthonormal set. So, one can think of the collection of factors as an *orthonormal set*, that is, a set of uncorrelated factors, whose scores are standardized so that the highest possible factor loading has the value 1.

Up to as many factors as there are items may be extracted. Such an extraction, however, would not provide useful generalizations to groups of similar constructs. For example, if each of the 12 galaxy statements under consideration loaded onto a unique factor, then there would be 12 groups, each with one galaxy statement, which would provide no practical insight as to how to group misconceptions together. On the other hand, one could extract just a single factor from the data. If all 12 galaxy statements loaded onto the same factor, however, then the analysis would not unveil correlated subsets from which students build a complete picture of galaxies. A single factor would instead mean that if a student endorses any one misconception about galaxies, then the student is likely to endorse all of the other misconceptions. For the purpose of identifying unique constructs within the group, an ideal goal is to extract an appropriate number of factors, which would certainly be fewer than the number of items in the analysis. For example, “Visual Properties” may represent one construct and thus one factor, whereas “Being in the Center” may represent another construct and thus a second factor. The goal of the analysis at this stage is to determine just how many factors to extract, a process which will be discussed in Section 6.3.

Once the extraction has taken place, the correlation between each item and each factor can be computed. Factor loadings provide relative measures for how strongly each item loads on the individual factors. Conceptually, if one represents each factor as a vector, then the factor loadings represent the displacement along each vector. In terms of the ABI data, factor loadings describe how closely each misconception relates to the construct of interest. For example, if “Visual Properties” and “Being in the Center” represent two orthogonal factors, then item sA220, the misconception that “all galaxies are spiral,” is expected to have a high loading on the “Visual Properties” factor and a low loading on the “Being in the Center” factor.

6.1.2 Rotation of factors

Once factor extraction has taken place, and factor loadings have been determined, the factor set should be transformed to provide the most insightful representation of the factor loadings. The transformation of the factor set is called *rotation*, as the factors can be thought of as dimensions in the vector sense. If the factors are to be kept uncorrelated with each other, then one performs an *orthogonal rotation*. As Lee and Ashton (2007) note, there are several kinds of orthogonal rotations. One such rotation is *varimax* rotation (Kaiser, 1958), in which the coefficients of the linear combination of standardized scores are chosen so that the item loadings on each factor are either very strong or near zero. In *quartimax* rotation (Neuhaus & Wrigley, 1954), the coefficients are chosen so that each variable loads as much as possible on a single factor. Rotation is the last process that one must perform to determine the final set of all factor loadings. The resulting set of all factor loadings from this process is the *factor structure*.

For the ABI data set, varimax rotation is ideal, because I am interested in examining groups of misconceptions that inter-correlate most strongly with each other. Quartimax rotation, on the other hand, is not ideal, because quartimax rotation places a high emphasis on forcing items to load on a single construct. The resulting factor structure from forcing items to load on a single construct may misrepresent the tendency for misconceptions to share multiple constructs. For example, sA225, “there are only a few galaxies,” may be associated with the “Visual Properties” construct (see discussion in Section 9.1), but also perhaps with the construct labeled “Spatial Distribution.” For the purposes of teaching, instructors can teach in the context of that misconception when the instructor covers galaxy appearances, and again when the instructor covers galaxy distributions. Since quartimax rotation underestimates the degree to which a misconception shares multiple constructs, the instructor who uses factor loadings derived from quartimax rotation may incorrectly conclude that misconceptions are more strictly compartmentalized into discrete groups than they really are.

6.2 Principal components analysis

In factor analysis, the factors represent a set of unmeasured variables that have been reduced from the original data set. The global process of determining just how many of these unmeasured variables succinctly and accurately represent the original data set is the holy grail of factor analysis. For the purposes of the ABI, each factor represents an overarching construct that groups together misconceptions. An analysis using the total variance of the items produces *principal components*, or factors whose eigenvalues produce relative measures of how much variation each factor has relative to the others. Factor analysis that proceeds in this way is called principal components analysis (PCA). In order to develop an accurate representative factor structure of the ABI topics, I have elected to use PCA. The components would then represent the latent constructs, e.g., “Visual Properties.” For simplicity, throughout the remainder of this chapter, I will refer to all of the factor analyses as simply factor analysis, and the components as factors that are generated from performing PCA.

6.3 Determining the number of factors to extract

In factor analysis, one is primarily concerned with creating a representative set of variables fewer than the number of items with which one started. At some point during factor analysis, one must determine just how many factors to extract. As Lee and Ashton (2007) point out, extracting more factors means accounting for more variance in the data, but also increased complexity in the data analysis.

6.3.1 The eigenvalue difference criterion

Several methods exist to determine the appropriate number of factors to extract. In factor analysis, a scree test (Cattell, 1966) gives the relative eigenvalues for up to as many factors as there are items. In a scree plot, the sum of the eigenvalues equals the number of items in the analysis. An example scree plot, for the 12 galaxy statements, is presented in Figure 9.1. Cattell claims that the number of factors to extract depends on the first noteworthy “jump” in the magnitudes between two sequential eigenvalues. The logic is that the extraction of one additional factor provides only a marginal increase in the cumulative variance accounted for compared to the extraction of the preceding factor. The problem with the criterion by Cattell, however, is that it is subject to interpretation. Depending on the data, some scree plots may have no obvious “jump” in sequential eigenvalues (e.g., Figure 8.1), or there may be multiple possible “jumps.” The decision as to where to assign the best “jump” for such plots may be ultimately arbitrary.

6.3.2 The criterion of eigenvalues greater than unity

A less subjective criterion is to use a criterion employed by Kaiser (1960), to extract as many factors as there are numbers of eigenvalues greater than one (henceforth, “eigenvalue criterion”). Eigenvalues of less than one represent factors that account for less standardized variance than that of any of the items. The logic behind choosing the value of one as the eigenvalue cutoff is that factors that account for less than the standardized variance of the items do not help in reducing the data set to a smaller number of variables.

There are two primary flaws with the eigenvalue criterion. First, the number of factors to extract is strongly dependent on the number of items in the analysis. The expectation is that a subset of the items is expected to represent a single construct (e.g., “galaxy appearance”), yet the number of constructs must be represented exclusively by the number of eigenvalues greater than one. Second, a factor whose eigenvalue is only marginally higher or marginally lower than one may be unnecessarily included or excluded, insomuch as to disrupt the factor structure.

6.3.3 The parallel analysis test

One method for adjusting the eigenvalues correspondingly is to use a parallel analysis (PA) test, developed by Horn (1965). The goal of PA is to determine the eigenvalues that would result from sampling variability in the data set. The logic of PA is that one should extract as many factors as there are eigenvalues greater than those generated by sampling variability. *In parallel analysis, the number of eigenvalues that are lower than those from the observed correlation matrix equals the number of factors to extract.* To perform PA, a set of mean eigenvalues is first generated each from a correlation matrix of normally-distributed random data. The mock data set has the same sample size and number of variables as the original data set, except that the data for each item are random and follow a normal distribution, e.g., $N \sim (0,1)$. Each pair of items is, therefore, expected to be essentially uncorrelated, since the data are random. A correlation matrix based on the mock data set is then generated, and a set of eigenvalues are calculated.

The process of going from a mock data set of random normal variables to a set of eigenvalues is repeated a predetermined number of times, depending on the sample size and number of items. In their studies, for example, Glorfeld (1995) and Crawford et al. (2010) use 1000-5000 replications; however, Humphreys and Montanelli Jr. (1975) obtained a relatively consistent factor structure with as few as 40-50 replications. For each factor, the mean and standard deviation of the eigenvalues over all the repetitions is determined. Ledesma and Valero-Mora (2007) cite various studies that use PA and a cutoff of the distribution of the eigenvalues for each proposed factor at the 95th percentile. From these studies, Ledesma and Valero-Mora argue that PA, among the factor extraction methods, is “an appropriate method to determine the number of factors,” and that the method is the least sensitive to particular aspects of the data set.

The primary weakness of PA is that, given the distribution of eigenvalues for each factor, *the choice of percentile* ultimately determines the number of factors to extract. As examined by Glorfeld (1995), the extraction of too many factors may result from choosing eigenvalues at the 50th percentile, while a more accurate factor structure may result from using the 95th or 99th percentile. Sometimes, however, PA has the tendency to support the extraction of too few factors even if the percentile cutoff is reduced to the 50th percentile. For the example of the 12 galaxy misconceptions from the ABI (see discussion in Section 9.1), the eigenvalue criterion suggests that there are three unique galaxy constructs: “Visual Properties,” “Being in the Center,” and “Spatial Distribution.” For comparison, I performed PA on the 12 galaxy misconceptions with 5000 repetitions and with percentile cutoffs ranging from the 5th percentile to the 95th percentile. In all cases, PA produced a single factor, a result which is rather contrary to the three-factor solution produced from the eigenvalue criterion. PA suggests that only the first factor is unlikely to arise by chance. Yet a single-factor representation provides little insight on the conceptual composition of the complete set of galaxy subtopics.

6.3.4 The comparison data method

More recently, Ruscio and Roche (2012) developed a computational method, called the comparison data method, which generates data for an increasingly higher number of factors until the generated eigenvalues reasonably match the observed eigenvalues. In the comparison data method, data with a known factorial structure is generated using one factor and compared to the observed data. If the first generated eigenvalue is significantly different from the first observed eigenvalue, then data is generated for two factors. The eigenvalues between the data sets are compared for successively more factors until the eigenvalue sets match reasonably well. Ruscio and Roche state that the comparison data method is more likely than the PA test to extract the correct number of factors for their data.

The comparison data method, however, faces a similar limitation as the PA test. Analogous to PA, *the choice to significance* of agreement between generated and observed eigenvalues ultimately determines how many factors to extract. As with PA, one must assign a rule for the cutoff after which no further factors may be extracted. Interestingly, the comparison data analysis presented by Ruscio and Roche also shows that the comparison data method tends to extract *fewer* factors than the PA test. In the preceding discussion of PA, only one galaxy factor was extracted. The comparison data method may potentially exacerbate the issue of the extraction of too few factors already present in my use of PA on the ABI data.

The discussion thus far illustrates that multiple methodologies exist for determining the appropriate number of factors to extract. Fundamentally, the issue with factor analysis is that all the variables are generally correlated with each other to some extent. The researcher is thus left to decide how highly inter-correlated a set of statements has to be to establish it as a group. Very few data sets in practice result in an unambiguous factor structure determination. The discussion of PA, in Section 6.3.3, and the comparison data method, in Section 6.3.4, does not convey that comparisons with simulated data sets necessarily provide either the "correct" factor structure or a more practical factor structure for studying the misconceptions that are the most strongly inter-correlated.

The preceding discussion suggests that when looking at the inventory data one topic at a time, the unambiguous eigenvalue criterion may be the most useful. Relative to the PA and comparison data methods, the eigenvalue criterion is more likely to extract too many factors, rather than too few factors. Therefore, I have elected to base the majority of my factor analysis using PCA of topics of the ABI with an eigenvalue criterion of one. In summary, I have elected to perform PCA by following the steps described in Table 6.1. I use SPSS to perform the analysis.

1. Construct a correlation matrix
2. Derive eigenvalues from the correlation matrix
3. Perform PCA with varimax rotation
4. Extract as many factors as there are eigenvalues greater than one
5. Generate a table of factor loadings
6. Sort the factor loadings as described in Section 6.4
7. Establish groups of items with the highest inter-correlations
8. If necessary, run a parallel analysis test to examine those groups that are unlikely to arise by chance

Table 6.1. Steps to take in performing principal components analysis

In the event that a meaningful discussion of all items in a large set of items is not feasible, I can restrict my analysis to just those factors that are supported by PA, as the factors are not likely to result by chance. Given my discussion of PA, in Section 6.3.3, I have elected to run 500 replications when I perform PA, as other studies showed that anywhere between 50 and 5000 replications should suffice. From my PA tests on data from each individual topic of the ABI, I also found no change to the number of factors that arise beyond chance when I increase the number of replications from 500 to 5000. In the event that one or more misconceptions tend to load on multiple factors (or "mix"), it would be safer to distinguish between those misconceptions that unambiguously load on a single factor, and then, during instruction of each subtopic, teach in the context of those misconceptions that load on multiple factors. I use a script developed in R (Kabacoff, 2012) to perform parallel analysis.

6.4 Sorting the factored items

Once factor analysis has been performed, and the appropriate number of factors has been extracted, rotation may be performed, and representative factor loadings may be reported. The most informative way to report the loadings is to present a table with the factors as columns, decreasing in variance to the right, and the items as rows. The items that are grouped together by the highest loadings on each factor represent independent groups of items. The organization of factor loadings in this way is the last step in determining the factor structure.

The following example illustrates the sorting method. For example, in Table 9.3, the first seven rows represent the group of seven items that load the most strongly on the first factor, labeled "I." The items in the first group are sorted by loading, from highest to lowest. The following three rows of Table 9.3 represent the group of three items that load most strongly on the second factor, labeled "II." The items in the second group are sorted in order of loading, from highest to lowest. The last two rows of Table 9.3 represent the group of two items that load most strongly on the third factor, labeled "III." The items in the third group are sorted by loading, from highest to lowest. Altogether, the process of sorting the items by their loadings on each factor produces a single table that can be used to represent the factor structure.

6.5 Application of PCA to an analysis of astrophysical data

In PCA, one can reduce the original variable set to a smaller set of variables. Often these variables can be grouped into two or more independent groups, such that the variables within each group are highly inter-correlated. The applications of PCA thus extend beyond the realm of the social sciences. For example, Einasto et al. (2011) used PCA to examine specific properties inherent to superclusters. Superclusters are systems of two or more interacting galaxy clusters and may be the largest known structures in the universe (Small, Ma, Sargent, & Hamilton, 1998, and references therein). In their study, Einasto et al. considered the size, shape, luminosity, peak density (the highest extent to which galaxies are clustered together), and galaxy member count of 125 superclusters in total, as well as the distances from the Earth to each of the superclusters in the sample. Together, these data were subjected to a PCA analysis. Einasto et al. found that the size, shape, luminosity, peak density, and galaxy member count of superclusters are all inter-correlated. These variables load on the first factor, which means that the data account for the largest proportion of the variance. This result just means that the values of the physical properties of the superclusters in the sample span evenly across a relatively broad range. Einasto et al. further showed that the variable for supercluster distances from the Earth loads most strongly on the second factor, and relatively weakly on the first factor, which means "the physical parameters of superclusters are not correlated with distance" (p. 4). These results were further supported by the results of additional tests conducted within the study. So, PCA may be used to analyze astrophysical data as well as data in astronomy involving human subjects.

CHAPTER 7

FACTOR STRUCTURE DEPENDENCE ON SAMPLE SELECTION

In Chapter 6, I presented an introduction to PCA as a method of grouping statements together from the ABI. The factor structure produced by PCA depends on the data within the correlation matrix, generated by the responses to statements within a particular topic of the ABI. The presentation of the factor structure depend on the choice of rotation (orthogonal vs. oblique) and is affected by variations in the corresponding correlation matrix. In this Chapter, I assess the dependence of the factor structure on the correlation matrices that arise from both semester-based and random selections of ABI responses.

7.1 Methodology of comparing factor structures

7.1.1 Challenges

The motivation to use PCA is to determine the best grouping of subtopics within each topic (e.g., galaxies, black holes) of the ABI. In order to justify the use of PCA, one should examine the dependence of the factor structure (as defined on page 118) on the sample responses. Of the six semesters under consideration, one might expect that the ABI data from the Fall 2009 to Fall 2012 semesters are expected to have the highest inter-semester consistency, because the course for each of the four semesters was taught by the same instructor using the same teaching pedagogy, and the format of the ABI statements did not change. As I showed in Chapter 5, however, the data are subject to selection effects, in that overall ABI scores varied significantly among the four semesters. The extent to which the variation in these scores affects correlations between statements in the ABI, however, requires that one inspect the data for each specific analysis. Hence, one cannot rule out that each of the four semesters may produce significantly different factor structures. Similarly, the extent of the similarity of the factor structures derived from data for the Spring 2013 and Fall 2013 semesters, compared to those for the Fall 2009 to Fall 2012 semesters likewise requires that one inspect the data for these particular cases. The inter-semester misconception score variations implicitly suggest that multiple alternate “best ways to teach” astronomy in AST 109 are possible, where the “ways” of teaching depend on the grouping of misconceptions in the factor structure.

When producing factor structures for a particular topic on a per-semester basis, the use of the eigenvalue criterion, as discussed in Section 6.3.2, may give a different number of factors for each semester. To establish direct comparisons of factor structures between semesters, one could require the same number of factors to be extracted for each semester. The choice to extract a certain number of factors, however, would then seem arbitrary without *a priori* knowledge of the resulting factor structures. Consider, as a best case scenario, that within each topic, each of the first four semesters produced the same number of factors. Then the actual comparison of factor structures still remains subject to interpretation of (i) the loadings of the statements within the groups and (ii) the loadings of other statements in the factor structure. Hence, one is required to establish a quantifiable and more consistent measure with which to compare factor structures. Therefore, I wish to examine the extent of the reliability of the correlations within the data.

7.1.2 Comparison of correlation matrices

Factor structures ultimately depend on the correlation matrix of responses to a set of items within a particular analysis. One consistent approach to comparing the factor structures is to compare correlation matrices between any pair of semesters. The motivation for correlation matrix comparisons is as follows: Consider two different items X and Y, such that their correlation $C_{XY,i}$, using data from semester i, can be calculated using Equation (3.4). Using data from another semester j, if the correlation

$C_{XY,j}$ between these same two items is the same or nearly the same as $C_{XY,i}$, then items X and Y are very likely to load the highest on the same factor together, in the factor structure for either semester. In other words, if the correlation between any two items remains about the same from semester to semester, then the factor structures among the semesters should all be reasonably similar. On the other hand, if the correlations $C_{XY,i}$ and $C_{XY,j}$ are very different between any two items X and Y, then the factor structures for both semesters may also be different, because different groups of inter-correlated items altogether should arise in the factor structures.

Therefore, the *magnitude of the difference* between the correlation coefficients, $|C_{XY,i} - C_{XY,j}|$ can be used to provide a fair, quantitative measure of the difference in resulting factor structures. Using the difference $|C_{XY,i} - C_{XY,j}|$, one can construct a matrix of differences in correlations between each pair of items X and Y. The number of unique comparisons is $K(K-1)/2$, where K is the number of items under consideration. A straightforward comparison between correlation matrices can be made by calculating the *mean* of the magnitude of all the unique comparisons, which, for simplicity, will be denoted m_{ij} and is given by

$$m_{ij} = \frac{2}{K(K-1)} \sum_{X=1}^{K-1} \sum_{Y>X}^{K} \left| C_{XY,i} - C_{XY,j} \right|. \tag{7.1}$$

The more that two correlation matrices differ from each other, the larger the value of m_{ij}. A representative mean can then be calculated using the means from all comparisons, which, for simplicity, will be denoted M_s, and is given by

$$M_s = \frac{2}{S(S-1)} \sum_{i=1}^{S-1} \sum_{j>i}^{S} m_{ij}, \tag{7.2}$$

where S is the number of subsamples (e.g., semesters), and $S(S-1)/2$ is the number of unique subsample comparisons. The standard deviation about M_s, denoted σ_s, represents the standard deviation of all the m_{ij} for each unique pair of semesters i and

j, calculated in the usual way for standard deviations. In the subsequent two paragraphs, I provide an interpretation for values of M_s.

Conceptually, M_s represents the mean shift in the correlation between any two statements, as a function of the semesters being analyzed. When one performs an analysis using data from multiple semesters, M_s can be thought of as *a relative measure of the sample dependence of the factor structure on the semester being analyzed*. A perfectly stable factor structure satisfies the condition $M_s = 0$, since the data from each semester would have the same correlations and thus produce the same factor structure. In practice, however, such a condition is rarely, if ever, satisfied. Higher values of M_s correspond to higher sample dependence on the factor structure, which also decreases the reliability of the factor structure generated from any one semester. In a sense, these factor structures are "unstable," because they may change significantly from one semester to another and so may not be reliably generalized.

In order to decide if a factor structure is reliable enough, one must choose a value of M_s below which the factor structure from the sample may be considered "reasonably stable." If it is shown that any one subsample (e.g., semester) may not produce a reliable factor structure, then one may consider grouping samples together, recoding the data into fewer response codes, or both, in an effort to reduce M_s. For example, in my sample of 403 students from the Fall 2009 to Fall 2012 semesters, I show that the statement-weighted mean M_s for the entire ABI is 0.125, and that values of M_s, depending on alternate comparisons of interest, is in the range 0.10 – 0.13. As a guideline, then, I consider $M_s \lesssim 0.1$ to represent a sample whose factor structure is reasonably stable and thus is a fair representation of the groups of inter-correlated misconceptions for the overall sample. Such a criterion is not unreasonable, because it requires that the mean difference in correlations from subsample to subsample (e.g., semester to semester) to be less than about 0.1. To clarify, the value of M_s below which one deems a factor structure reliable is arbitrary and, therefore, is not intended to represent a rigorous criterion for my analysis. Given the criterion that $M_s \lesssim 0.1$ corresponds to a reasonably stable factor structure, values of $M_s \gg 0.1$ correspond to an "unstable" factor structure: one that does not fairly represent groups of inter-correlated misconceptions for the overall sample.

Of note, one cannot use M_s solely to predict specific changes to a given factor structure. M_s represents the mean difference in correlations between misconceptions and so does not provide any information about changes in *specific* correlations between statements. The actual factor loadings also depend on the rotation of the factors. So, $M_s \lesssim 0.1$ does *not* represent differences of factor loadings of less than about 0.1. The calculation of M_s goes one route, namely, by computing the mean difference in correlations; the determination of the factor structure goes an entirely different route, namely, by using the eigenvalues of the correlation matrix. Because factor loadings depend on particular details regarding the factor analysis methodology, there is no straightforward relationship between changes in M_s and changes to the factor structure. Moreover, M_s does not contain any information about which misconceptions become more or less statistically correlated between subsamples. Therefore, while M_s may be considered a useful quantitative measure of the difference in factor structures, M_s cannot predict (i) the magnitude of the changes in the factor loadings or (ii) specific re-groupings of misconceptions into new groups.

7.1.3 Random sampling of correlation matrices

An alternate way of assessing the dependence of the factor structure on the ABI responses is to generate correlation matrices from random samples (taken from the composite data set) of all responses from all semesters. With each random sample pair the correlation differences $|C_{XY,k} - C_{XY,l}|$ can be calculated, where k and l represent two different random samples. Using all of the random samples, a representative random-sample mean and representative standard deviation can be calculated, which, for simplicity, will be respectively denoted M_r and σ_r. The calculation of M_r is given by the expression in Equation (7.2), where, here, S represents the number of random samples. From the Fall 2009 to Fall 2012 semesters, for example, the sample size $n = 403$ is adequately large for one to perform thousands of comparisons, using a sufficiently large subset, e.g., of 100 subjects, since there are, on average, about 100 subjects per semester. Hence, if enough random samples are performed, then the mean of the correlation differences reasonably model a normal distribution.

One can then test for significance between M_r and M_s, with the null hypothesis that M_s does not differ significantly from M_r. In theory, M_r is expected to be lower than M_s, because M_r uses data from all semesters, whereas M_s is more subject to semester selection effects. Hence, one can use a one-tailed confidence interval (Lacey, 1998), in which one expects that all possible values for M_r should be lower than M_s, and that the probability of obtaining a value of M_r much lower than M_s tapers off (like a tail on either side of a normal distribution). In such a one-tailed distribution, M_s is "significantly" higher than M_r if $M_s > M_r + 1.64\sigma_r$, which is where $p = 0.05$. If the result is that M_s, which is determined using data from one semester at a time, is significantly different from M_r, then no one semester can be reliably generalized to represent the factor structure for all possible semesters.

In summary, I propose two methods for assessing the dependence of the factor structure on the ABI responses. The first method uses the magnitude of the differences in statement correlations between any two semesters. The representative mean and standard deviation of the mean differences within each semester comparison, are respectively denoted M_s and σ_s. The second method involves the random sampling of data from all semesters and the computation of the differences between pairs of correlation matrices derived from the random samples. The representative mean and standard deviation, of the mean differences within each random sample comparison, are respectively denoted M_r and σ_r.

7.2 Correlation matrix comparisons for the ABI

7.2.1 Fall 2009 to Fall 2012

Using the Fall 2009 to Fall 2012 data ($n = 403$), I constructed 100 correlation matrices, each using data from 100 subjects selected at random, for each section of the ABI. For each section, I calculated both representative means and standard deviations, i.e., M_r, σ_r, M_s, and σ_s, along with the z-score that represents the number of standard deviations in σ_r that M_s differs from M_r. A z-score of 1.64 or higher represents a significant difference.

Table 7.1 presents, for each topic, representative means and standard deviations, the z-score, and whether or not the random sample gives a significantly more reliable factor structure than the factor structure generated by a single semester. The number of analyzed statements associated with each topic is included in parentheses. The z-scores are calculated based on the values of M_r and σ_r, prior to rounding. Note that the section labeled Three Planets combines all statements pertaining to the Venus, Mars, and Saturn topics from the ABI (refer to Appendix A).

Table 7.1 shows that for a substantial portion of the ABI, random samples are significantly better than the use of a single semester to generalize student responses to the ABI. Note that the z-scores are also lower for those sections of the ABI with fewer statements, which is to be expected, because fewer statements means more statistical power and, hence, higher consistency of correlations between semesters. The mean M_s for all sections of the ABI, weighted by the number of analyzed statements in each section, is 0.125.

7.2.2 Estimating mean correlation shifts using a subset of all ABI statements

To check that the mean M_s for all sections should be around 0.125, I considered 30 statements, selected at random from the entire ABI of 215 statements.

	M_r	σ_r	M_s	σ_s	z	**Significant?**
Stars (34)	0.1009	0.0066	0.1262	0.0135	3.81	Yes
Solar System (30)	0.1006	0.0072	0.1200	0.0105	2.69	Yes
Moon (24)	0.1023	0.0082	0.1328	0.0152	3.72	Yes
Three Planets (17)	0.1042	0.0109	0.1321	0.0147	2.56	Yes
Earth (37)	0.1015	0.0072	0.1229	0.0069	2.96	Yes
Sun (32)	0.1013	0.0074	0.1252	0.0074	3.22	Yes
Galaxies (12)	0.1004	0.0131	0.1347	0.0235	2.62	Yes
Black Holes (13)	0.1032	0.0131	0.1207	0.0107	1.34	No
General Astrophysics (16)	0.1022	0.0106	0.1161	0.0080	1.31	No

Table 7.1. Factor structure stability, per topic, for the Fall 2009 to Fall 2012 semesters

Using these 30 statements, I calculated M_s using data from the Fall 2009 to Fall 2012 semesters in otherwise the same manner as with the analysis in Section 7.2.1. From these 30 statements, I find $M_s = 0.131$, $\sigma_s = 0.008$. Hence, an adequately-sized random sample of statements is sufficient at quickly estimating M_s for the whole inventory.

My results thus far suggest that the mean inter-semester difference in correlations between any statement pair is expected to be around 0.125, which is somewhat high compared to M_r. Hence, no particular semester is likely to give a reliable representation of the factor structure. Notably, the data come from students who have voluntarily responded to the ABI. Hence, the presence of selection effects require that data from more than one semester be included in factor analysis. The consequence is that *data from a single semester may not produce a reliable factor structure*.

7.2.3 Spring 2013 to Fall 2013

To assess the effect of teaching pedagogy on statement correlations, I determined M_s using data from the Spring 2013 and Fall 2013 semesters ($n = 236$), for both of which I employed the revised ABI format (refer to Section 2.3). Since there is only one comparison, I do not report σ_s. Using the same 30 statements selected at random, I find $M_s = 0.110$, which is significantly lower than the quoted result of 0.131 using data from the Fall 2009 to Fall 2012 semesters. For comparison, the analysis in Section 3.5.5 shows that, between the Spring 2013 and Fall 2013 semesters, the overall fraction of misconceptions endorsed after instruction does not differ significantly between the two semesters. So, one should not be surprised that M_s is lower for the comparison between the Spring 2013 and Fall 2013 semesters than for the first four semesters together.

7.2.4 Comparing semester groups between the old and new ABI formats

Thus far in Section 7.2, I have shown that statement correlations among the first four semesters tend to vary by about 0.13, and that statement correlations between the Spring 2013 and Fall 2013 semesters tend to vary by about 0.11. The uncertainties in these estimates are each on the order of 0.01. Since these two semester groups use different ABI formats, one may ask whether or not the ABI format affects the nature of correlations between statements, as 86 of the 215 statements were rephrased as true in the Spring 2013 and Fall 2013 semesters.

To examine the extent to which the ABI formats influence statement correlations, from the same set of 30 statements, I aggregated data from the Fall 2009 to Fall 2012 semesters together as one group with 403 subjects. Then I aggregated data from the Spring 2013 and Fall 2013 semesters together as a second group with 236 subjects. I then determined M_s between the two groups. For the full sample size of $N = 639$, I find $M_s = 0.097$. Note that by aggregating the two semester groups together, my overall sample size increases by approximately one third. When I randomly select and remove one third of subjects from each group, I find $M_s = 0.108$. Both results are *lower* than those quoted for within each semester group, is consistent with the hypothesis that *correlations between statements are more significantly affected by inter-semester selection effects than by the rephrasing of statements on an inventory of misconceptions*.

In Section 5.1, I explained that the Fall 2009 to Fall 2012 semesters suffered from significant variability with regard to the fraction of misconceptions endorsed even after instruction in AST 109. The selection effects left open the question of whether or not one could perform a reliable correlation analysis of the affected statements, which, in turn, left open the question of whether one could determine groups of correlated misconceptions. Having explicitly quantified the differences in correlation matrices, in which M_s ranges from 0.10 to 0.13 and with an uncertainty of 0.01, regardless of the analyses that I have considered, I can now confidently conclude that *despite significant inter-semester variability of the fraction of misconceptions endorsed after instruction, correlations between misconceptions are not significantly affected*.

7.3 Implications for a principal components analysis of the ABI

In summary, I have created a method to quantify the difference between two factor structures, by using data from their respective parent correlation matrices. When I analyze the differences in correlation matrices between semesters vs. within random samples taken from all semesters, I get M_r significantly less than M_s, which means that data taken from only a single semester cannot produce a reliable factor structure. In my global analysis of the ABI data, I wish to study the misconceptions that are most easily dispelled during one's childhood or adolescence. Therefore, I have elected to utilize data from at least the Fall 2009 to Fall 2012 semesters.

Ideally, one would like to use data from as many semesters as possible to perform PCA, especially since the format of the ABI does not seem to present any statistical concern with regard to misconceptions correlations among all six semesters (refer to Section 7.2.4). The inclusion of data from the aggregate of both ABI formats, however, presents a new issue. The groups of statements that result from performing PCA no longer represent *correlated misconceptions*. Instead, the resulting factor analysis would produce groups of "sometimes misconceptions" items. For example, one group may contain five statements, three of which are worded as misconceptions, and two of which were phrased as true statements for just the Spring 2013 and Fall 2013 semesters. Another group may have always-false statements. Still yet another group may have statements, all of which were worded as true for the two semesters. These groups no longer have the same latent context of *misconception* endorsement. As I explain in Section 13.3, my analysis of these "sometimes misconceptions" items using data from all six semesters would raise concerns regarding the dimensionality of my data. The issue of increased dimensionality is further compounded by the data in the Spring 2013 semester, in which the instructor and teaching pedagogy were different altogether compared to those of the other five semesters. *These issues are circumvented entirely by using data from exclusively the first four semesters, in which statements in the ABI are consistently phrased as misconceptions.* Therefore, I will concern my analysis for the remainder of my dissertation with data from exclusively the Fall 2009 to Fall 2012 semesters.

CHAPTER 8

COMPARATIVE PLANETOLOGY - OR ONE PLANET AT A TIME?

In Chapter 6, I presented an introduction to PCA as a method of grouping misconceptions together within a topic. The process that I use to perform PCA is summarized Table 6.1. In Chapters 8-12, I apply PCA to the data from various sections of the inventory. The goal of Chapters 8-12 is to examine the groups of misconceptions that result from performing PCA. In this Chapter, I perform PCA on 17 planet misconceptions pertaining to Venus, Mars, and Saturn, and I discuss the results in the context of which method for teaching planets is more effective at dispelling the incorrect beliefs of the students.

Table 8.1 lists the planet statements, presented to the students for the Fall 2009 to Fall 2012 semester administrations of the ABI. The following five statements in the ABI pertain to misconceptions about Venus: sA106, sA107, sA108, sA109, and sA110. The following eight statements in the ABI pertain to misconceptions about Mars: sA161, sA164, sA165, sA166, sA167, sA168, sA169, and sA170. The following four statements in the ABI pertain to misconceptions about Saturn: sA171, sA172, sA174, and sA174.

8.1 Comparative planetology

Our solar system is divided into four inner planets with rocky or rocky and liquid water surfaces (Mercury, Venus, Earth, Mars), two gas giants (Jupiter, Saturn), and two ice giants (Uranus, Neptune). Astronomy textbooks typically discuss that all of the planets in our solar system have a common origin: a collapsing protosolar cloud. That the planets in the solar system originated from a common protosolar cloud has motivated many astronomy textbook authors to write about the planets by comparing

	Planet	Statement
sA106	Venus	life as we know it can exist on Venus
sA107	Venus	clouds on Venus are composed of water, like clouds on earth
sA108	Venus	Venus is very different from earth in size
sA109	Venus	Venus is a lot like the earth in temperature
sA110	Venus	Venus is always the first star out at night
sA161	Mars	Mars is green (from plant life)
sA164	Mars	Mars has running water now
sA165	Mars	Mars could be made inhabitable
sA166	Mars	Mars is the second largest planet
sA167	Mars	life, when it did exist on Mars, was quite advanced
sA168	Mars	there are Lowellian canals on Mars built by intelligent beings
sA169	Mars	Mars is Hot because it is red ... Mars – god of fire
sA170	Mars	Mars is the sister planet to earth in physical properties and dimensions
sA171	Saturn	Saturn is the only planet with rings
sA172	Saturn	Saturn's rings are solid
sA174	Saturn	Saturn's rings are caused by the planet spinning so fast
sA176	Saturn	Saturn has only one ring

Table 8.1. Planet statements as presented to the students in the Fall 2009 to Fall 2012 semesters

their compositional properties, one layer at a time (F. W. Taylor, 2011). The "comparison" approach is formally known as *comparative planetology*.

With the aid of spacecraft imagery of planet atmospheres and surfaces, a plethora of textbooks geared toward students taking a course in introductory astronomy now teach the planets using the comparative planetology paradigm. Rather than treat the planets, one *planet* at a time, many such textbooks are written so as to construct the layers of multiple planets at a time, one *layer* at a time (Beatty, Petersen, & Chaikin, 1999; Bennett, Donahue, Schneider, & Volt, 2012; Fix, 2004; Freedman, Geller, & Kaufmann III, 2011; Hartmann, 1999; Hester et al., 2007; Lewis, 2004; Palen, Lay, Smith, & Blumenthal, 2011; Schneider & Arny, 2007; Seeds, 2010; Shipman, 1978; Shu, 1982; Slater & Freedman, 2012; Snow, 1990; Snow & Brownsberger, 1997; Zeilik, 1997, 2002)In contrast, many introductory astronomy textbooks present each planet, one at a time (Arny, 2004, 2006; Chaisson & McMillan, 1996; Comins &

Kaufmann III, 2012; Comins, 2013; Dixon, 1992; Engelbrektson, 1994; Kaler, 1994; Kaufmann III & Freedman, 2002; Koupelis & Kuhn, 1999; Koupelis, 2014; Pasachoff, 1998; Pasachoff & Filippenko, 2007). As F. W. Taylor (2011) claims, comparative planetology helps the students to build a "unified climate model" for Venus, Earth, and Mars, and that such a model is "a more reliable representation of all three planets" (p. 898). Svedham (2011) adds that such a unified model provides the students with "an improved understanding of the evolution of each of these three planets" (p. 887).

Despite the recent growing trend of teaching through comparative planetology, no definitive studies have been published thus far to support either method of teaching the planets over the other. As of August 21, 2013, out of all the publications listed on the Center for Astronomy Education website (2013), and out of all papers ever published by Astronomy Education Review (2013), not even one resource that I found compared the effectiveness of the comparative planetology paradigm vs. the instruction of one planet at a time, on either scientific learning or the unlearning of incorrect beliefs associated with the planets.

8.2 Principal components analysis on the planet statements

I performed PCA on the ABI data for the planets from the Fall 2009 to Fall 2012 semesters, each taught under NFC. I determined the mean misconception retainment score for the four semesters ($M = 1.85$, $SD = 0.30$). The internal consistency of the ABI responses, given by coefficient alpha, was $\alpha = 0.80$, which is considered adequate (Schmitt, 1996). For the purpose of performing PCA, missing cases were excluded listwise, which is to say for a particular student if no response is provided for one or more items in the topic, then all of the student's responses to the remaining items are ignored. The remaining sample size is 398.

Table 8.2 presents a summary of planet statement correlations, taken from the Fall 2009 to Fall 2012 semesters. Since the correlation matrix is its own transpose, the correlations in Table 8.2 are sufficient to reproduce the correlation matrix. (The column labels are not to be confused with correlations.)

The scree plot, in Figure 8.1, shows the distribution of eigenvalues for the planet statements, taken from the Fall 2009 to Fall 2012 semesters. The eigenvalues of the first six factors are 4.19, 1.71, 1.33, 1.21, 1.02, and 0.88. Using the eigenvalue criterion, PCA supports five groups of planet statements from the extraction of five factors, which cumulatively account for 56% of the total variance.

Table 8.3 presents the factor loadings for the planet statements after extraction of five factors. The factors are labeled as columns in order of decreasing variance from left to right, and the statements are labeled as rows. The rows have been further labeled with a symbol corresponding to a statement pertaining to Venus (♀), Mars (♂), or Saturn (♄). Statements with the highest loadings on a particular factor are grouped together and labeled in **bold**. Loadings of magnitude 0.3 or higher of statements outside their primary groups on another factor account for at least 9% of the variance, which, for the purpose of discussing other noteworthy factor loadings, I consider significant, and so are marked with an asterisk. A significant loading on another factor means that students are likely to unlearn the misconception associated with the statement when instructors teach to it along with the group of highest-loading statements on the same factor. Loadings of less than 0.3 are not marked with an asterisk and represent misconceptions that are unlikely to be unlearned when teaching to the statement associated with it and to those in the group.

The factor structure presented in Table 8.3 suggests that students tend to unlearn most planet misconceptions one planet at a time, though some specific planet misconceptions are better unlearned through comparisons. In the first factor, five statements, all about Mars, load the most strongly, so the first factor can be labeled "Mars." The first three of these statements, sA161, sA167, and sA168, pertain to life on Mars. In the second factor, three statements, all about Saturn, load the most

sA:	106	107	108	109	110	161	164	165	166	167	168	169	170	171	172	174
107	0.333															
108	0.182	0.256														
109	0.328	0.444	0.330													
110	0.120	0.229	0.262	0.190												
161	0.221	0.118	0.179	0.137	0.139											
164	0.230	0.160	0.100	0.200	0.170	0.236										
165	0.242	0.149	0.126	0.218	0.072	0.100	0.302									
166	0.147	0.181	0.315	0.280	0.188	0.371	0.246	0.174								
167	0.179	0.047	0.163	0.130	0.129	0.413	0.291	0.243	0.304							
168	0.132	0.033	0.170	0.110	0.122	0.412	0.318	0.189	0.242	0.534						
169	0.122	0.090	0.135	0.132	0.113	0.310	0.263	0.145	0.269	0.351	0.275					
170	0.031	0.134	0.282	0.087	0.215	0.096	0.112	0.159	0.248	0.106	0.097	0.214				
171	0.191	0.235	0.213	0.153	0.112	0.110	0.183	0.103	0.260	0.168	0.067	0.272	0.254			
172	0.150	0.267	0.168	0.224	0.086	0.093	0.174	0.147	0.207	0.194	0.070	0.212	0.159	0.470		
174	0.091	0.071	0.158	0.148	0.126	0.151	0.209	0.208	0.272	0.151	0.184	0.145	0.177	0.218	0.296	
176	0.214	0.283	0.230	0.181	0.067	0.201	0.241	0.176	0.291	0.096	0.090	0.300	0.090	0.284	0.397	0.228

Table 8.2. Correlation matrix of responses to the planet statements

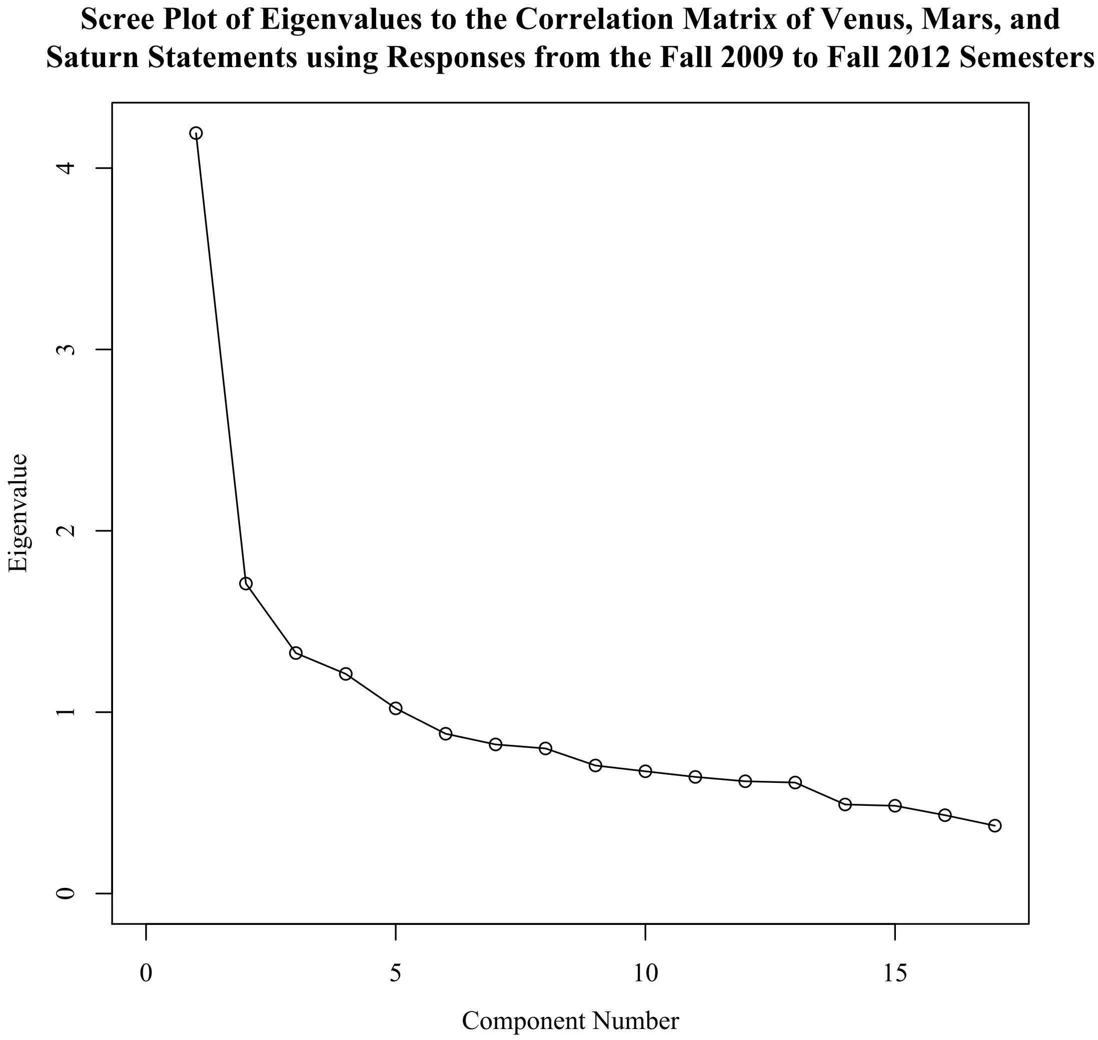


Figure 8.1. Scree plot of eigenvalues from the correlation matrix of responses to the planet statements

	I	II	III	IV	V
sA161 (♂)	**0.753**	0.080	0.161	0.088	–0.090
sA167 (♂)	**0.746**	0.066	0.025	0.058	0.219
sA168 (♂)	**0.743**	–0.076	0.007	0.085	0.240
sA169 (♂)	**0.534**	0.413*	–0.032	0.087	0.027
sA166 (♂)	**0.425**	0.293	0.137	0.402*	0.059
sA172 (♄)	0.017	**0.764**	0.138	0.043	0.150
sA171 (♄)	0.064	**0.716**	0.070	0.192	0.044
sA176 (♄)	0.151	**0.653**	0.275	–0.038	0.060
sA107 (♀)	–0.050	0.240	**0.722**	0.202	–0.015
sA109 (♀)	0.057	0.089	**0.700**	0.243	0.133
sA106 (♀)	0.177	0.088	**0.678**	–0.082	0.191
sA170 (♂)	0.018	0.209	–0.165	**0.706**	0.185
sA110 (♀)	0.103	–0.090	0.217	**0.633**	0.047
sA108 (♀)	0.145	0.132	0.317*	**0.624**	–0.058
sA165 (♂)	0.079	0.019	0.210	0.048	**0.797**
sA164 (♂)	0.359*	0.134	0.196	0.002	**0.543**
sA174 (♄)	0.074	0.367*	–0.105	0.240	**0.503**

Table 8.3. Planet statement factor loadings

strongly, so the second factor can be labeled “Saturn.” In the third factor, three statements, all about Venus, load the most strongly, so the third factor can be labeled “Venus.”

The fourth and fifth factors consist of mixed planet statements. In the fourth factor, statements sA108, “Venus is very different from earth in size,” and sA170, “Mars is the sister planet to earth in physical properties and dimensions,” represent comparisons to the Earth. Interestingly, statement sA110, “Venus is always the first star out at night,” loads on the same factor, although sA110 bears little resemblance to comparisons of either planet with the Earth. To a lesser extent, statement sA166, “Mars is the second largest planet,” loads on the same factor. Therefore, the fourth factor can be labeled “Mixed Earth.”

The fifth factor contains of two Mars statements regarding the surfaces of Mars: sA164, “Mars could be made inhabitable,” and sA165, “Mars has running water now.” The last item in the factor, sA174, “Saturn’s rings are caused by the planet spinning so fast," continues the underlying theme of features that move or wrap around the surfaces of the planets. The fifth factor thus consists of misconceptions regarding the visual features around these worlds. Therefore, the fifth factor can be labeled "Mixed Visual Features."

A total of six out of the original 17 planet statements have higher loadings on one of the "Mixed" factors than their own representative planet factor. For example, statement sA164 loads significantly on the "Mars" factor, but its loading is highest on the "Mixed Surfaces" factor, which suggests that the misconception associated with sA164 is more effectively unlearned by showing comparisons about surface features around the planets than by teaching about Mars, independent of the other planets. Also, two statements about Venus, sA108 and sA110, have their highest loadings on the "Mixed Earth" factor, rather than the "Venus" factor. Table 8.4 summarizes the factor labels for the Planets statements.

Factor	**Label**
I	Mars
II	Venus
III	Saturn
IV	Mixed Earth
V	Mixed Visual Features

Table 8.4. Factor labels for the planets statements

Together, the "Mixed" planet factors contain six out of the original 17 planet statements, the misconceptions of which are more effectively unlearned by planet comparisons than by teaching each planet strictly one at a time. The remaining 11 misconceptions are best unlearned by teaching the properties of the associated planets, one planet at a time. Hence, my results neither convey that one should teach exclusively one planet at a time, nor convey that planet misconceptions are best untaught through comparative planetology. As a reminder to the reader, comparative planetology involves the conceptual construction of planet *layers*, rather than the construction of each planet separately. *My results suggest that the planet* *misconceptions may be unlearned most effectively, first, by teaching each planet with its own features, and then by making targeted comparisons between their surface features and the relation of the planets to the Earth.*

CHAPTER 9

PRINCIPAL COMPONENTS ANALYSIS ON GALAXY AND BLACK HOLE MISCONCEPTIONS

In this Chapter, I discuss the PCA results on the ABI data for the Galaxies and Black Holes topics of the ABI. The Galaxies and Black Holes topics contain the fewest number of misconceptions (12 and 13, respectively), so the statistical power (that is, the ratio of subjects to items) of my analysis for these topics is the highest relative to the statistical power of all other topics of the ABI. My presentation of the PCA results in this Chapter is similar to that in Chapter 8, in which I use ABI data from only the Fall 2009 to Fall 2012 semesters, each taught under NFC. I then compare the resulting factor structures to those using data from the Spring 2013 and Fall 2013 semesters, in which the format of the ABI changed, with the rephrasing of the 87 items from false to true, and the order of statement presentation was randomized (refer to Section 2.3)

9.1 Galaxies

9.1.1 Fall 2009 to Fall 2012

sA218. the Milky Way is the only galaxy
sA219. the solar system is not in the Milky Way (or any other) galaxy
sA220. all galaxies are spiral
sA221. the Milky Way is the center of the universe
sA222. the Sun is at the center of the Milky Way galaxy
sA224. the Sun is at the center of the universe
sA225. there are only a few galaxies
sA226. the galaxies are randomly distributed
sA227. we see all the stars that are in the Milky Way
sA228. all galaxies are the same in size and shape
sA230. the Milky Way is just stars — no gas and dust
sA231. new planets and stars don’t form today

Table 9.1. Galaxy statements as presented to the students in the Fall 2009 to Fall 2012 semesters

I performed PCA on the ABI data for the galaxy statements from the Fall 2009 to Fall 2012 semesters. Table 9.1 lists the galaxy statements, presented to the students for the Fall 2009 to Fall 2012 administrations of the ABI. I determined the mean misconception retainment score for the four semesters ($M = 1.78$, $SD = 0.33$). The internal consistency of the ABI responses was $\alpha = 0.81$, which, as discussed in Section 3.4, is considered adequate (Schmitt, 1996). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 392. Table 9.2 presents a summary of galaxy statement correlations.

The scree plot, in Figure 9.1, shows the distribution of eigenvalues for the galaxy statements. The eigenvalues of the first four factors are 3.99, 1.13, 1.07, and 0.86. Using the eigenvalue criterion, PCA supports three groups of galaxy statements from the extraction of three factors, which cumulatively account for 52% of the total variance.

Table 9.3 presents the factor loadings for the galaxy statements after extraction of three factors. The factors are labeled as columns in order of decreasing variance from left to right, and the statements are labeled as rows. Statements with the highest loadings on a particular factor are grouped together and labeled in **bold**. Factor loadings of 0.3 or higher on statements outside of their primary groups are marked with an asterisk.

The factor structure presented in Table 9.3 suggests that students may unlearn galaxy misconceptions best when galaxies are taught in subtopics about: their visual properties; our *not* being in the center of the galaxy or the universe; and, spatial distributions, namely, those of (i) galaxies on the sky and (ii) the stars that we can see in our own galaxy. The hypothesis that these particular subgroups promote the most effective unlearning of galaxy misconceptions may be tested in future research. In Table 9.3, statements sA218, sA219, sA220, sA225, sA228, sA230, and sA231 are all grouped together and load the most on the first factor. The first five of these statements

	sA218	sA219	sA220	sA221	sA222	sA224	sA225	sA226	sA227	sA228	sA230
sA219	0.267										
sA220	0.286	0.236									
sA221	0.347	0.221	0.253								
sA222	0.292	0.163	0.206	0.470							
sA224	0.383	0.223	0.191	0.464	0.477						
sA225	0.348	0.271	0.323	0.367	0.325	0.395					
sA226	0.076	0.031	0.176	0.133	0.232	0.098	0.268				
sA227	0.283	0.108	0.194	0.244	0.196	0.213	0.289	0.237			
sA228	0.294	0.219	0.262	0.314	0.173	0.203	0.392	0.175	0.302		
sA230	0.282	0.246	0.242	0.259	0.227	0.350	0.374	0.174	0.270	0.376	
sA231	0.289	0.253	0.253	0.282	0.293	0.305	0.279	0.117	0.227	0.310	0.395

Table 9.2. Correlation matrix of responses to the galaxy statements

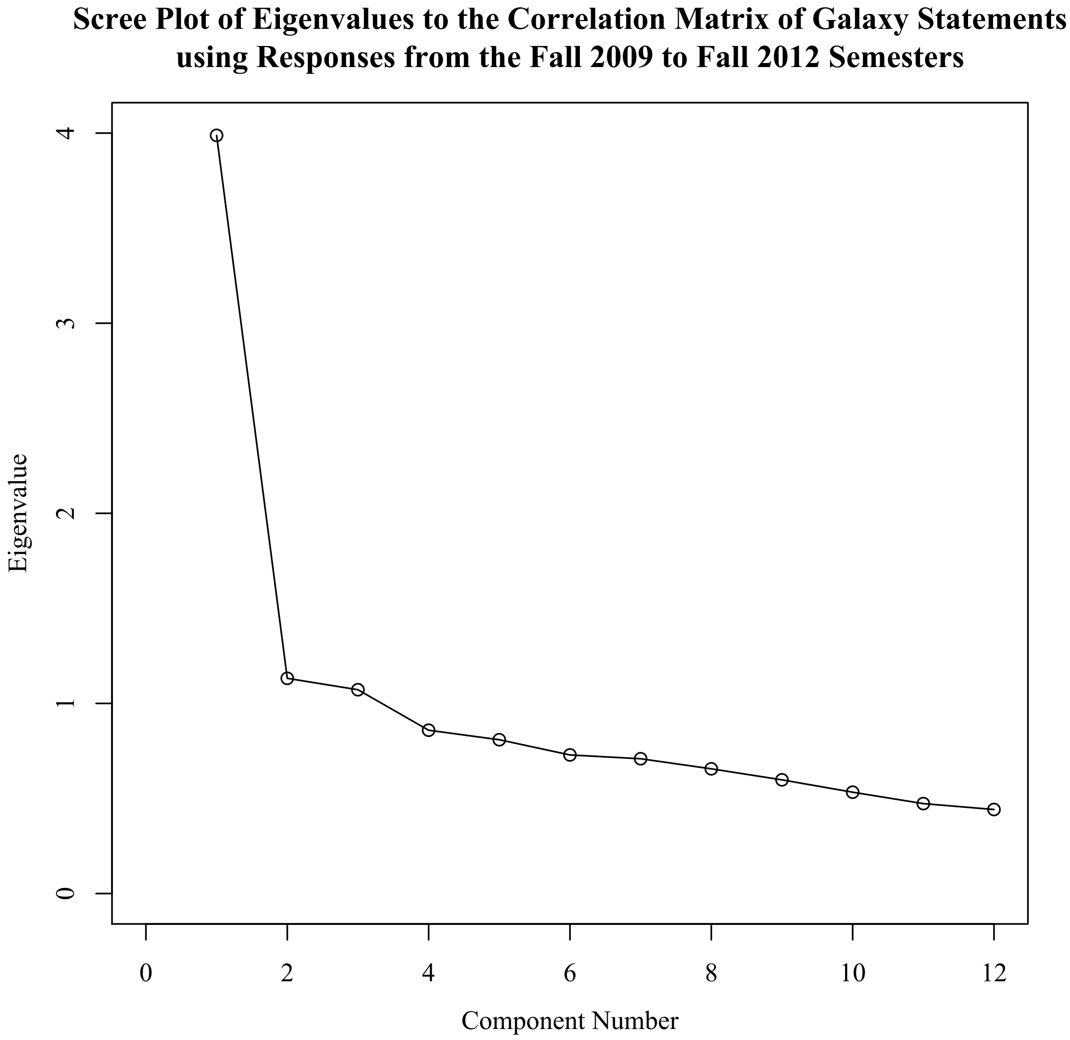


Figure 9.1. Scree plot of eigenvalues from the correlation matrix of responses to the galaxy statements

	I	II	III
sA228	**0.645**	0.038	0.316*
sA219	**0.638**	0.140	–0.261
sA230	**0.623**	0.175	0.206
sA231	**0.572**	0.264	0.054
sA220	**0.527**	0.113	0.203
sA218	**0.502**	0.422*	–0.003
sA225	**0.491**	0.362*	0.349*
sA222	0.047	**0.805**	0.206
sA224	0.244	**0.773**	0.003
sA221	0.253	**0.711**	0.107
sA226	0.004	0.105	**0.831**
sA227	0.342*	0.124	**0.545**

Table 9.3. Galaxy statement factor loadings

pertain to the variety of galaxy types, whereas the last two relate to visual properties about the solar system location and the formation of planets and stars. Overall, these statements pertain to general visual characteristics about galaxies, so the first factor can be labeled “Visual Properties.”

In the second factor, statements sA221, “the Milky Way is the center of the universe,” and sA224, “the Sun is at the center of the universe,” pertain to the misconception of being at the center of the universe. Statement sA222, “the Sun is at the center of the Milky Way galaxy,” pertains to the misconception of being at the center of the Milky Way galaxy. Overall, these statements belong to the group in the second factor which can be labeled “Being in the Center.”

In the third factor, statement sA226, “the galaxies are randomly distributed,” loads the highest and pertains to spatial distribution of galaxies. Statement sA227, “we see all the stars that are in the Milky Way,” pertains to the spatial distribution of the stars in the Milky Way and is grouped together with sA226. Together, these statements belong to the third group in the factor labeled “Spatial Distribution.” Table 9.4 summarizes the factor labels for the galaxies statements.

Factor	**Label**
I	Visual Properties
II	Being in the Center
III	Spatial Distribution

Table 9.4. Factor labels for the galaxies statement groups

Some items have significant loadings on multiple factors. For instance, statement sA225, “there are only a few galaxies,” tends to load on all three factors. Hence, given any of the three groups of most inter-correlated statements, there is some tendency for students also to unlearn or retain the misconception associated with sA225. Statement sA219, “the solar system is not in the Milky Way (or any other) galaxy,” has a non-significant loading on the second factor, which pertains to “Being in the Center,” but has a marginally significant *negative* loading on the third factor. The negative loading implies that whether students unlearn or endorse the misconceptions associated with sA226 or sA227, the *opposite* is somewhat likely to occur with sA219. Statement sA219 also has a relatively low loading on the second factor, which suggests that *the tendency to abandon misconceptions about being in the center of the galaxy or the center of the universe is separate from the tendency to abandon the misconception that the solar system is not in our galaxy.* In other words, from a conceptual standpoint, there is no significant correlation between “being in the center” and “not being in it altogether.”

9.1.2 Spring 2013 and Fall 2013

In Section 2.2, I outlined two unique formats of the ABI. The first format of the ABI consisted of topics that were presented essentially in the same order as the material in the course. The modified format of the ABI was presented to the students in the Spring 2013 and Fall 2013 semesters, with the order of the statements randomized. In Section 7.2.4, I stated that the mean difference in statement correlations from one semester group to another should change by about 0.1. Given that there are 12 galaxy statements under consideration, one may thus expect some variations in the factor structure when data from the Spring 2013 and Fall 2013 semesters are used. Table 9.5 lists the galaxy statements as presented to the students in the Spring 2013 and Fall 2013 semesters, in which eight of the 12 galaxy statements (marked with an "n" in their respective labels) were changed from false to true and then presented to the students.

sA218n.	the Milky Way is one of many galaxies
sA219n.	the solar system is inside the Milky Way galaxy
sA220.	all galaxies are spiral
sA221.	the Milky Way is the center of the universe
sA222n.	the Sun is far away from the center of the Milky Way galaxy
sA224.	the Sun is at the center of the universe
sA225n.	there are billions of galaxies
sA226.	the galaxies are randomly distributed
sA227n.	we can see only some of the stars in the Milky Way
sA228n.	galaxies have a variety of sizes and shapes
sA230n.	the Milky Way contains gas and dust
sA231n.	new planets and stars are forming today

Table 9.5. Galaxy statements as presented to the students in the Spring 2013 and Fall 2013 semesters

Following the same data reduction as described in Section 9.1.1, I performed PCA on the ABI data for the galaxy statements, using data from the Spring 2013 and Fall 2013 semesters. The extraction of three factors, cumulatively accounting for 47% of the variance, was supported by the Kaiser criterion. Table 9.6 presents the factor loadings for the galaxy statements after extraction of three factors.

The data in Table 9.6 show the following:

- Statements sA221, sA222n, and sA224, which are associated with the theme of "Being in the Center," under the column for the second factor in Table 9.4, now load the highest on the *first* factor. The statements themselves remain members of the same group of the most inter-correlated items.

	I	II	III
sA219n	**0.730**	0.029	–0.093
sA224	**0.711**	0.175	0.179
sA221	**0.660**	0.193	0.029
sA218n	**0.656**	0.097	–0.084
sA222n	**0.502**	0.181	0.276
sA230n	0.098	**0.652**	0.101
sA228n	0.086	**0.645**	0.071
sA227n	0.006	**0.597**	–0.006
sA225n	0.340*	**0.524**	–0.149
sA220	0.312*	**0.515**	0.227
sA231n	0.129	**0.510**	–0.036
sA226	0.018	0.025	**0.934**

Table 9.6. Galaxy statement factor loadings using data from the Spring 2013 and Fall 2013 semesters

- Statements sA220, sA225n, sA228n, sA230n, and sA231n, which are associated with the theme of "Visual Properties," under the column for the first factor in Table 9.4, now load the highest on the *second* factor. The statements themselves remain members of the same group of the most inter-correlated items.

- Statement sA226, which is associated with the theme of "Spatial Distribution," under the column for the third factor in Table 9.4, still loads the highest on the third factor.

- Statement sA227n, "we can see only some of the stars in the Milky Way," whose counterpart (sA227) is a member of the group under "Spatial Distribution," now shares its highest inter-correlations with members of the group under "Visual Properties." The context of sA227n, as read by the students, is likely to prompt visual recall of what they can see when they think of the Milky Way galaxy, so it is not surprising to see sA227n grouped together with other statements about the visual properties of galaxies.

- Statements sA218n and sA219n, whose counterparts (respectively sA218 and sA219) are members of the group under the "Visual Properties" factor, are now the most inter-correlated with statements associated with the theme of "Being in the Center." The context of sA218n, "the Milky Way is one of many galaxies," does support that students may be referring back to their sense of direction, in that students may think about whether or not they can look to find other galaxies other than our own. One's sense of direction is also utilized in order for students to decide if we are at the center of the galaxy or the universe. Likewise, the context of sA219n, "the solar system is inside the Milky Way galaxy," supports the same utilization of sense of location. So, although sA218n and sA219n load the most strongly with a new group of statements compared to before the statements were rephrased, the regrouping is not so serious as to require any reclassification of the factor structure.

In this section, I have compared the factor structure of the 12 galaxy statements, using two independent data subsets (Fall 2009 to Fall 2012 vs. Spring 2013 and Fall 2013), each of adequate statistical power, from the overall set of six semesters. *My comparisons reveal that the factor structures that result from performing PCA on each subset, conceptually speaking, have undergone only minor changes, a result which is consistent with the expectation that I set in Section 7.2.4.* One particular noteworthy change is that factors I and II have been essentially switched, though the statements within the corresponding groups have undergone only minor changes. The reordering of factors merely reflects that the data in the new first factor accounts for more variance than the data in the new second factor. Hence, the galaxy statement factor structure is not significantly affected by (i) the phrasing of the statements as false vs. true or (ii) the order of statement presentation.

sA232. black holes create themselves from nothing
sA233. black holes last forever
sA234. black holes really don't exist
sA235. black holes are empty space
sA237. black holes do not have mass
sA238. black holes are like huge vacuum cleaners, sucking things in
sA240. black holes are doors to other dimensions
sA242. black holes can be seen visually, like seeing a star or planet
sA243. we could live in a voyage through a black hole
sA244. we could travel through time in a black hole
sA245. black holes get bigger forever and nothing can stop them from doing so
sA246. black holes are actual holes in space
sA247. a single black hole will eventually suck in all the matter in the universe

Table 9.7. Black hole statements as presented to the students in the Fall 2009 to Fall 2012 semesters

9.2 Black holes

9.2.1 Fall 2009 to Fall 2012

I performed PCA on the ABI data for the black hole statements from the Fall 2009 to Fall 2012 semesters. Missing cases were excluded listwise. Table 9.7 lists the 13 black hole statements, presented to the students for the Fall 2009 to Fall 2012 administrations of the ABI. I determined the mean misconception retainment score for the four semesters ($M = 1.89$, $SD = 0.35$). The internal consistency of the ABI responses was $\alpha = 0.80$, which, as discussed in Section 3.4, is considered adequate. For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 398. Table 9.8 presents a summary of black hole statement correlations, taken from the Fall 2009 to Fall 2012 semesters.

The scree plot, in Figure 9.2, shows the distribution of eigenvalues for the black hole statements. The eigenvalues of the first four factors are 3.96, 1.27, 1.11, and 0.92. Using the eigenvalue criterion, PCA supports three groups of black hole statements from the extraction of three factors, which cumulatively account for 49% of the total variance.

	sA232	sA233	sA234	sA235	sA237	sA238	sA240	sA242	sA243	sA244	sA245	sA246
sA233	0.371											
sA234	0.142	0.052										
sA235	0.346	0.271	0.267									
sA237	0.267	0.243	0.209	0.426								
sA238	0.171	0.191	0.169	0.294	0.272							
sA240	0.188	0.149	0.168	0.145	0.110	0.129						
sA242	0.215	0.191	0.076	0.185	0.221	0.156	0.188					
sA243	0.216	0.251	0.230	0.265	0.208	0.147	0.267	0.298				
sA244	0.167	0.183	0.181	0.233	0.124	0.157	0.458	0.191	0.363			
sA245	0.154	0.370	0.094	0.294	0.202	0.185	0.213	0.336	0.310	0.268		
sA246	0.332	0.266	0.159	0.452	0.317	0.234	0.270	0.297	0.367	0.267	0.364	
sA247	0.270	0.326	0.169	0.222	0.264	0.276	0.291	0.159	0.296	0.294	0.332	0.380

Table 9.8. Correlation matrix of responses to the black hole statements

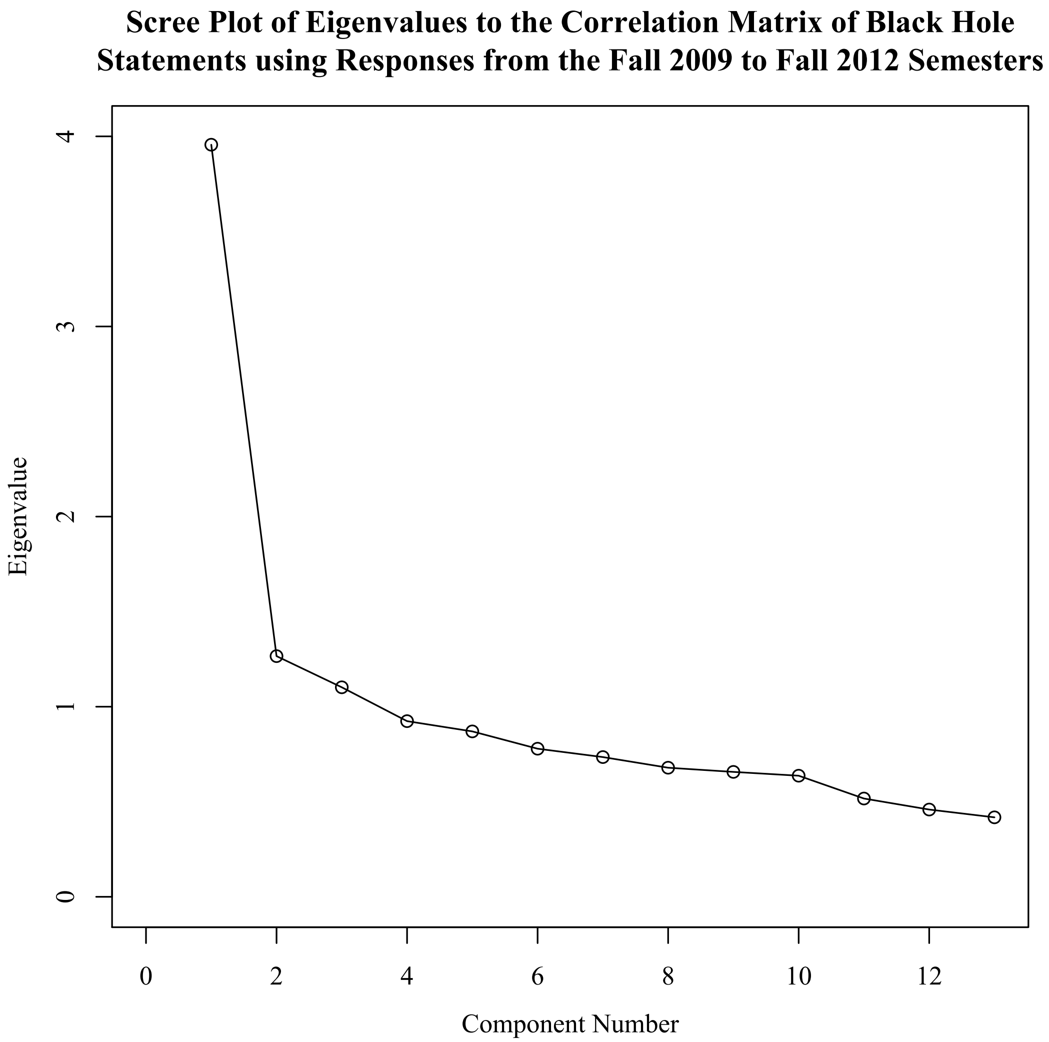


Figure 9.2. Scree plot of eigenvalues from the correlation matrix of responses to the black hole statements

Table 9.9 presents the factor loadings for the black hole statements after extraction of three factors. The factors are labeled as columns in order of decreasing variance from left to right, and the statements are labeled as rows. Statements with the highest loadings on a particular factor are grouped together and labeled in **bold**. Factor loadings of 0.3 or higher on statements outside of their primary groups are marked with an asterisk.

	I	II	III
sA235	**0.713**	0.243	0.105
sA237	**0.691**	0.225	–0.019
sA234	**0.588**	–0.328*	0.400*
sA238	**0.553**	0.122	0.092
sA232	**0.436**	0.426*	0.042
sA233	0.215	**0.687**	0.016
sA245	0.097	**0.656**	0.268
sA242	0.081	**0.541**	0.211
sA246	0.426*	**0.463**	0.293
sA247	0.291	**0.399**	0.375*
sA244	0.070	0.153	**0.774**
sA240	0.023	0.123	**0.762**
sA243	0.201	0.329*	**0.536**

Table 9.9. Black hole statement factor loadings

The factor structure presented in Table 9.9 suggests that *students tend to unlearn black hole misconceptions most efficiently when black holes are taught in subtopics about: their physical properties; their appearance and fate; and their relation to science fiction.* In Table 9.9, statements sA232, sA234, sA235, sA237, and sA238 are all grouped together and load the most on the first factor. These statements pertain to the existence of black holes and their physical properties, i.e., if they have mass, and if they are "empty space." Statement sA246, "black holes are actual holes in space," also has a very significant loading on the first factor, and, like the other statements on the first factor also pertains to a physical property of black holes. Therefore, the first column can be labeled "Black Hole Properties."

In the second column, statements sA233, sA242, sA245, sA246, and sA247 load the most strongly on the second factor. Statements sA233, "black holes last forever," and sA245, "black holes get bigger forever and nothing can stop them from doing so," have distinctly the highest loadings on the factor and pertain to the fate of black holes, as do sA232 and sA247. Statement sA246 pertains to the ability to see black holes. Overall, the second factor seems to have a mixture of two themes related to both the appearance and the fate of black holes, and so the second factor can be labeled "Appearance and Fate." Note that the negative loading of statement sA234, "black holes really don't exist," on the second factor is significant. The negative loading is not surprising, since students who endorse misconceptions about the appearance and fate of black holes are naturally likely to expect black holes to exist in the first place.

In the third factor, the three statements that load the highest on it, sA240, sA243, and sA244, all correspond to ideas perpetrated by science fiction. The ideas that black holes permit time or space travel through them or to other dimensions are entirely fictional in nature. Statement sA247, "a single black hole will eventually suck in all the matter in the universe," loads significantly on both the second and third factors. Such loadings may not be surprising, since the second factor pertains to the fate of black holes, and the third factor pertains to fictional tales of black holes. Statement sA234 also has a positive, significant loading on the third factor, which says that students who endorse misconceptions about traveling through black holes naturally expect black holes to exist. The third column can be labeled "Science Fiction." Table 9.4 summarizes the factor labels for the black holes statements.

Factor	Label
I	Black Hole Properties
II	Appearance and Fate
III	Science Fiction

Table 9.10. Factor labels for the black holes statement groups

9.2.2 Spring 2013 and Fall 2013

In Section 9.1.2, I compared galaxy statement factor structures, using data taken from the Fall 2009 to Fall 2012 semesters as one subset and data taken from the Spring 2013 and Fall 2013 semesters as another subset. The format of the ABI underwent significant changes between Fall 2012 and Spring 2013; yet my analysis in Section 9.1.2 showed that these changes had only little impact on the galaxy statement factor structure. To test whether or not changes to the ABI format may affect the factor structure of statements pertaining to other topics of the ABI, I performed the same set of analytical comparisons using the 13 black hole statements. Table 9.11 presents the black hole statements as presented to the students in the Spring 2013 and Fall 2013 semesters, in which six of the 13 black hole statements (marked with an “n” in their respective labels) were changed from false to true and then presented to the students.

sA232.	black holes create themselves from nothing
sA233n.	black holes have finite lifetimes
sA234n.	black holes are real objects in space
sA235.	black holes are empty space
sA237n.	black holes have mass
sA238.	black holes are like huge vacuum cleaners, sucking things in
sA240n.	black holes are not really doors to other dimensions
sA242n.	black holes are invisible
sA243n.	a voyage through a black hole would be fatal
sA244.	we could travel through time in a black hole
sA245.	black holes get bigger forever and nothing can stop them from doing so
sA246.	black holes are actual holes in space
sA247.	a single black hole will eventually suck in all the matter in the universe

Table 9.11. Black hole statements as presented to the students in the Spring 2013 and Fall 2013 semesters

Following the same data reduction as described in Section 9.2.1, I performed PCA on the ABI data for the black hole statements, using data from the Spring 2013 and Fall 2013 semesters. The extraction of three factors, cumulatively accounting for 42% of the variance, was supported by the Kaiser criterion. Table 9.12 presents the factor loadings for the black hole statements after extraction of three factors.

	I	II	III
sA237n	**0.700**	–0.136	–0.114
sA232	**0.662**	0.127	0.210
sA235	**0.604**	0.337*	0.074
sA243n	**0.557**	0.205	0.051
sA234n	**0.481**	0.327*	0.027
sA242n	**0.318**	–0.260	0.228
sA240n	0.080	**0.695**	–0.166
sA244	0.063	**0.685**	0.093
sA246	0.267	**0.548**	0.178
sA245	–0.001	0.046	**0.765**
sA247	0.318*	0.365*	**0.470**
sA238	0.187	–0.220	**0.448**
sA233n	0.326*	–0.217	**–0.426**

Table 9.12. Black hole statement factor loadings using data from the Spring 2013 and Fall 2013 semesters

The data in Table 9.12 show the following:

- Statements sA232, sA234n, sA235, and sA237n, which are associated with the theme of "Black Hole Properties," under the column for the first factor in Table 9.10, remain members of the same group and still load the highest on the first factor. Statement sA238 now loads the most strongly on the third factor.
- Statements sA242n, "black holes are invisible," and sA243n, "a voyage through a black hole would be fatal," have been added to the first group. The context of sA242n, as read by the students, is likely to prompt recall of the properties of black holes as would the other statements in the group, so it is not surprising to see sA242n grouped together with other statements about the properties of black holes. One could also assert that when prompted to read sA243n, students may assert that the risk of fatality when traveling through a black hole is a characteristic property of black holes. Hence, the overall theme of the first factor still remains consistent with that of "Black Hole Properties." In other words, the regrouping is not so serious as to require any reclassification of the first factor.

- Statements sA233n, sA245, and sA247, which are associated with the theme of "Appearance and Fate," under the column for the second factor in Table 9.10, now load the highest on the *third* factor. The statements themselves remain members of the same group of the most inter-correlated items.

- Statements sA240n and sA244, which are associated with the theme of "Science Fiction," under the column for the third factor in Table 9.10, now load the highest on the *second* factor. Missing from the group is sA243n, which, as previously discussed, now loads the most strongly on the first factor. Added to the new second group, with sA240n and sA244, is sA246, "black holes are actual holes in space." Since statements sA240n and sA244 are associated with misconceptions pertaining to travel through time and/or to other dimensions, one should not be too surprised to see sA246 included in the group, since one first ought to construct an image of a "black hole" in order to make any sense of what it would mean to go through one.

In this section, I have compared the factor structure of the 13 black hole statements, using two independent data subsets, each of adequate statistical power, from the overall set of six semesters. The comparison was performed in a similar manner to that in Section 9.1.2. My comparisons reveal that the factor structures that result from performing PCA on each subset, conceptually speaking, underwent only minor changes, a result which is consistent with the expectation that I set in Section 7.2.4. One particular noteworthy change is that factors II and III have been essentially switched, which again merely reflects that the data in the new first factor accounts for more variance than the data in the new second factor. Hence, the black hole statement factor structure is not significantly affected by (i) the phrasing of the statements as false vs. true or (ii) the order of statement presentation. These are the same conclusions which I drew in Section 9.1.2, which suggests that the modifications to the

ABI produce only minor changes to the factor structure of statements for each of the remaining topics of the ABI.

CHAPTER 10

PRINCIPAL COMPONENTS ANALYSIS ON THE OTHER INVENTORY SECTIONS

In this Chapter, I continue my PCA of the ABI sections. The titles of the remaining sections of the ABI are: Stars, Solar System, Moon, Earth, Sun, and General Astrophysics (e.g., cosmic rays, telescopes, cosmology). Respectively, the number of statements in each of these sections are: 34, 30, 24, 37, 32, and 16. All sections except for General Astrophysics contain at least 24 statements, so the presentation of a table consisting of statement correlations for all the statements in a topic (e.g., Table 8.2) would be very large and so will be omitted for each of the remaining topics. Also, the statistical power (that is, the sample size to the number of statements) is reduced due to the high number of statements in these topics. For the Fall 2009 to Fall 2012 semesters, the statistical power of the Moon statements is $403/24 = 16.8$, and the statistical power of the Earth statements is $403/37 = 10.9$, compared to $403/12 = 33.6$ for galaxies and $403/13 = 31.0$ for black holes.

10.1 Reduced representation of the factor structure

To present a reasonably compact analysis of the remaining topics, I will present reduced factor structure tables, one per topic, instead of tables of correlations among all the items. The reduced factor structure tables will consist of exclusively those groups of items with the highest loadings on a particular factor. The groups are again identified in **bold**. For each factor, loadings of 0.3 or higher in magnitude on another factor will also be presented in order of decreasing loading below each group. All other factor loadings will be omitted. The reader may refer to Appendix A for a list of statements for each of the remaining sections of the ABI.

10.2 Stars

I performed PCA on the ABI data for 34 statements about stars from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 34 statements about stars. I determined the mean misconception retainment score for the four semesters ($M = 1.79$, $SD = 0.28$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 394. The internal consistency of the ABI responses was $\alpha = 0.88$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.1 shows the distribution of eigenvalues for statements about stars. Using the eigenvalue criterion, PCA supports eight groups of statements about stars from the extraction of eight factors, which cumulatively account for 49% of the total variance. The first factor accounts for 21% of the total variance, while the remaining factors each account for up to 5.3% of the total variance. At the 50th percentile, extraction of the first three factors is consistent with the results of performing parallel analysis on the ABI data, which says that the first three groups are unlikely to arise by chance.

Table 10.1 presents a reduced representation of factor loadings (as discussed in Section 10.1) for the statements about stars after extraction of eight factors. The factors are sequenced in order of decreasing variance. Statements with the highest loadings on a particular factor are grouped together and labeled in **bold**. Additional factor loadings of 0.3 or higher are included and sequenced in order of highest to lowest loadings below their respective factor groups. All other factor loadings have been omitted.

In order of decreasing loading, the group of statements in the first factor consists of sA34, “stars are fixed in space,” sA33, “we see the same constellations at night throughout the year,” sA22, “all stars are stationary — fixed on the celestial sphere,” sA20, “stars just existed — they don’t make energy or change size or color,”

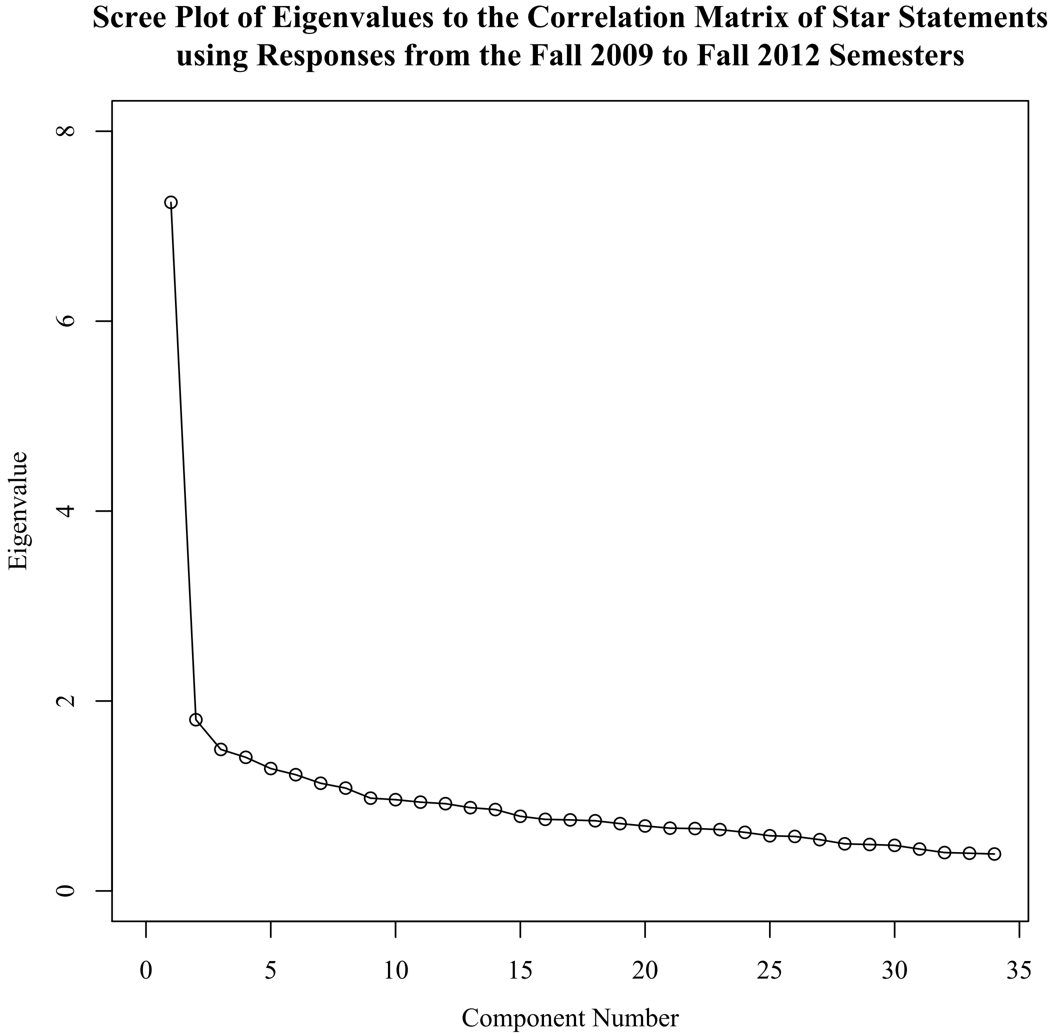


Figure 10.1. Scree plot of eigenvalues from the correlation matrix of responses to statements about stars

I		II		III		IV	
sA34	**0.713**	**sA17**	**0.586**	**sA32**	**0.664**	**sA35**	**0.683**
sA33	**0.611**	**sA11**	**0.564**	**sA14**	**0.598**	**sA28**	**0.621**
sA22	**0.552**	**sA18**	**0.526**	**sA3**	**0.587**	**sA21**	**0.562**
sA20	**0.547**	**sA6**	**0.524**	**sA15**	**0.445**	**sA37**	**0.425**
sA23	**0.413**	**sA7**	**0.520**	**sA30**	**0.443**	**sA31**	**0.395**
sA31	0.375	sA15	0.408	sA24	0.368	sA30	0.334
sA1	0.352	sA20	0.350	sA25	0.340		
sA4	0.328			sA18	0.304		

V		VI		VII		VIII	
sA10	**0.699**	**sA27**	**0.566**	**sA12**	**0.620**	**sA2**	**0.727**
sA13	**0.551**	**sA29**	**0.539**	**sA16**	**0.571**	**sA25**	**0.553**
sA9	**0.512**	**sA24**	**0.493**	**sA4**	**0.400**	**sA5**	**0.537**
sA1	**0.375**	sA23	0.380	**sA8**	**0.383**	sA8	0.378
sA14	0.332	sA22	0.366	sA7	0.323	sA7	0.309
sA15	0.301	sA5	0.355	sA5	0.318		

Table 10.1. Reduced representation of factor loadings of statements about stars

and sA23, “stars emit only one color of light.” To a lesser, extent, sA31, ““metals” have always existed in the universe,” sA1, “all of the stars were created at the same time,” and sA4, “all stars are white,” also load on the first factor. These statements pertain to the nature of stars as being fixed points of light, as if they don’t change. Hence, the first group consists of statements where stars are characterized as unchanging points of light.

A brief discussion of the second and third groups follows. The statements in the second group are sA17, “all stars are smaller than the Sun,” sA11, “stars last forever,” sA18, “the galaxy, solar system and universe are the same things,” sA6, “we are looking at stars as they are now,” and sA7, “stars actually twinkle — change brightness.” To a lesser extent, sA15, “all stars have same color and size,” and sA20 also load on the second factor. The second group seems to consist mostly of misconceptions about how stars actually appear relative to each other.

The statements in the third group are sA32, “stars follow you in your car,” sA14, “all stars are the same distance from the Earth,” sA3, “all of the stars are about as far away from the Earth as the Moon," sA15, "all stars have same color and size," and sA30, "stars run on fuel: gasoline or natural gas." To a lesser extent, sA24, "stars are closer to us than the Sun," sA25, "there are exactly 12 constellations," and sA18 load on the third factor. The third group seems to consist mostly of statements about tracking the stars in the sky.

A brief thematic interpretation of the remaining groups follow. The statements in the fourth group are sA35, "stars in a binary system (two stars bound together by their gravity) would quickly collide," sA28, "a nova is the most powerful explosion," sA21, "all stars end up as white dwarves," sA37, "all stars are isolated from all other stars (none are binary)," and sA31, " 'metals' have always existed in the universe." These statements are associated with misconceptions about binary systems and star deaths. The statements in the fifth group are sA10, "all stars have planets," sA13, "all stars are evenly distributed on the celestial sphere," sA9, "stars have spokes," and sA1, "all of the stars were created at the same time." These statements are associated with misconceptions about the initial formation of stars. The statements in the sixth group are sA27, "all the stars in an asterism move together," sA29, "stars in the Milky Way are as close to each other as planets are to the Sun," and sA24, "stars are closer to us than the Sun." These statements are associated with incorrect spatial reasoning regarding the distances to the stars. The statements in the seventh group are sA12, "the brighter a star is, the hotter it is," sA16, "pulsars are pulsating stars," sA4, "all stars are white," and sA8, "the north star is the brightest star in the sky." These statements are associated with misconceptions about the brightness of stars. The statements in the eighth group are sA2, "there are 12 zodiac constellations," sA25, "there are exactly 12 constellations," and sA5, "the constellations are only the stars we connect to make patterns." These statements are associated with misconceptions about constellations.

10.3 Solar system

I performed PCA on the ABI data for the 30 statements about the solar system from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 30 statements about the solar system. I determined the mean misconception retainment score for the four semesters ($M = 1.81$, $SD = 0.29$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 398. The internal consistency of the ABI responses was $\alpha = 0.86$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.2 shows the distribution of eigenvalues for statements about the solar system. Using the eigenvalue criterion, PCA supports nine groups of statements about the solar system from the extraction of nine factors, which cumulatively account for 54% of the total variance. The first factor accounts for 20% of the total variance, while the remaining factors account each account for up to 6.3% of the total variance. At the 50th percentile, extraction of the first two factors is consistent with the results of performing PA on the ABI data. Table 10.2 presents a reduced representation of factor loadings for the solar system statements after extraction of nine factors.

In order of decreasing loading, the group of statements in the first factor consists of sA42, "comet tails are burning — because the comet is moving so fast," sA41, "Mercury is so named because there is much mercury on it," sA40, "the asteroid belt is an area like we see in Star wars, very densely packed," sA60, "all stars have prograde rotation (spin same way as the Earth)," and sA45, "a shooting star is actually a star whizzing across the universe or falling through the sky." Statements sA43, "there is plant life on other planets in our solar system," sA56, "comets last forever," and sA48, "the Solar System is the whole universe or the whole galaxy," also load significantly on the first factor. The statements in the first group seem to pertain to identifications of solar system objects. Hence, the first group consists of statements about solar system object identifications.

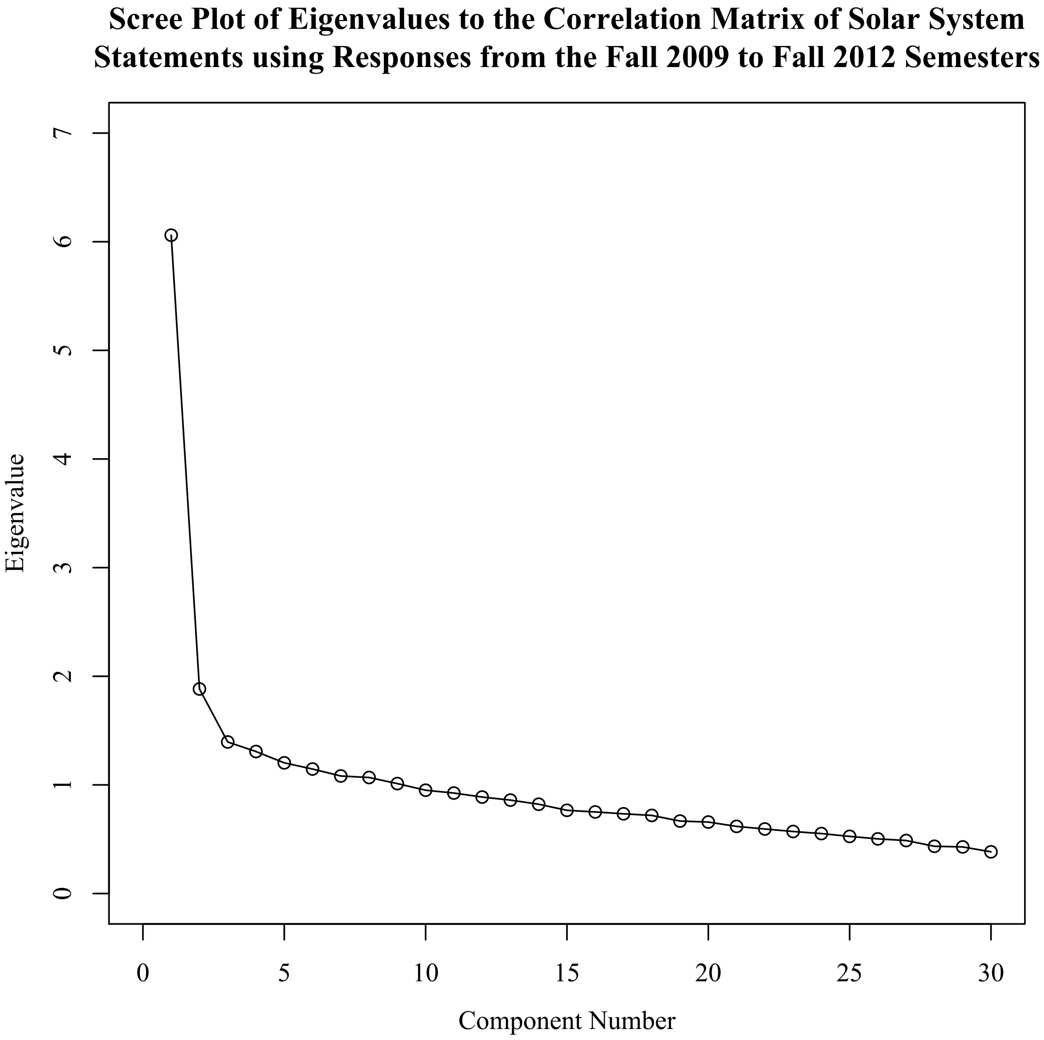


Figure 10.2. Scree plot of eigenvalues from the correlation matrix of responses to statements about the solar system

I		II		III		IV		V	
sA42	**0.578**	**sA68**	**0.657**	**sA47**	**0.693**	**sA58**	**0.661**	**sA53**	**0.687**
sA41	**0.566**	**sA67**	**0.640**	**sA49**	**0.658**	**sA44**	**0.615**	**sA51**	**0.660**
sA40	**0.526**	**sA56**	**0.462**	**sA76**	**0.435**	**sA46**	**0.544**	**sA43**	**0.432**
sA60	**0.494**	**sA48**	**0.378**	**sA72**	**0.344**	**sA54**	**0.399**	sA76	0.373
sA45	**0.489**	sA66	0.327	**sA50**	**0.313**	sA40	0.332	sA54	0.363
sA43	0.356	sA72	0.305	sA48	0.336				
sA56	0.340			sA52	0.321				
sA48	0.326								

VI		VII		VIII		IX	
sA75	**0.699**	**sA62**	**0.720**	**sA59**	**0.707**	**sA52**	**0.610**
sA70	**0.652**	**sA66**	**0.514**	**sA57**	**0.581**	**sA69**	**0.602**
sA45	0.362	**sA63**	**0.513**	sA56	0.350	sA60	0.454
sA44	0.330	sA43	–0.341				
sA72	0.327						

Table 10.2. Reduced representation of factor loadings of statements about the solar system

The statements in the second group are sA68, “we do not have telescopes in space,” sA67, “humans have never landed a spacecraft on another planet,” sA56, “comets last forever,” and sA48, “the Solar System is the whole universe or the whole galaxy,” whose loading is much weaker than those of the preceding three statements. Statements sA66, “optical telescopes are the only “eyes” astronomers have on the universe,” and sA72, “there are many galaxies in a solar system,” also load significantly on the second factor. The second group can be considered a factor about exploration of solar system bodies.

A brief thematic interpretation of the remaining groups follow. The statements in the third group are sA47, “the asteroid belt is between Earth and Mars,” sA49, “Jupiter is almost large and massive enough to be a star,” sA76, “Jupiter’s great red spot is a volcano erupting,” sA72, “there are many galaxies in a solar system,” and sA50, “all orbits around Sun are circular.” Other than the correlation between incorrect statements regarding Jupiter, there does not seem to be any particular theme for the statements in this group. The statements in the fourth group are sA58, “Mercury (closest planet to the Sun) is hot everywhere on its surface," sA44, "Pluto is always farther from the Sun than is Neptune," sA46, "Jovian planets (Jupiter, Saturn, Uranus, Neptune) have solid surfaces," and sA54, "all constellations look like things they are named." These statements may be associated with one's misinterpretation of the surface textures of planets and the appearance of constellations, since the planets are often presented in the media as colorful rigid circles, and constellations are often presented in the media as stars connected in ways to represent mythological creatures. There does not otherwise appear to be any particular thematic interpretation for why these statements are grouped together.

The statements in the fifth group are sA53, "Pluto is a large, jovian (Jupiter-like) planet," sA51, "planets revolve around the Earth," and sA43, "there is plant life on other planets in our solar system." There does not seem to be any particular theme for the statements in this group; however, these statements may be partly associated with misconceptions about other planets having plant life, like the Earth, or being large like Jupiter. Strangely enough, statements sA44 and sA46, from the fourth group, which are associated with misconceptions about planets in the outer solar system, do not load significantly on the fifth factor, which suggests that misconceptions regarding the visual presentation of the planets and constellations in the media are generally not correlated with misconceptions about the relative sizes or orbits of the planets.

A thematic interpretation of the remaining groups continues. The statements in the sixth group are sA75, "comets are solid, rocky debris," and sA70, "comets are molten rock hurtling through space at high speeds and their tails are jet wash behind them." These statements are associated with misconceptions about comets. The statements in the seventh group are sA62, "there are no differences between meteors, meteorites, meteoroids," sA66, "optical telescopes are the only 'eyes' astronomers have on the universe," and sA63, "asteroids, meteoroids, comets are same." That these statements are grouped together suggests that one who studies astronomy may become quickly confused when one does not sit down to acknowledge the fundamental differences among meteors, meteorites, meteoroids, and asteroids. The statements in the eighth group are sA59, “the day on each planet is 24 hours long,” and sA57, “each planet has one moon.” These statements are associated with the misconception that the concept of a ‘day’ should be similar to that on Earth. The statements in the ninth group are sA52, “all planets orbit exactly in the plane of the ecliptic,” and sA69, “all planets have been known for hundreds of years.” These statements are associated with the misconception that the orbital inclination of planets in our solar system is a fair representation of orbital inclination of planets in other “solar systems.”

10.4 Moon

I performed PCA on the ABI data for the 24 statements about the Moon from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 24 statements about the Moon. I determined the mean misconception retainment score for the four semesters ($M = 1.82$, $SD = 0.29$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 397. The internal consistency of the ABI responses was $\alpha = 0.84$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.3 shows the distribution of eigenvalues for statements about the Moon. Using the eigenvalue criterion, PCA supports five groups of statements about the Moon from the extraction of five factors, which cumulatively account for 45% of the total variance. The first factor accounts for 22% of the total variance, while the remaining factors account each account for up to 7.3% of the total variance. At the 50th percentile, extraction of the first two factors is consistent with the results of performing PA on the ABI data. Table 10.3 presents a reduced representation of factor loadings for the moon statements after extraction of five factors.

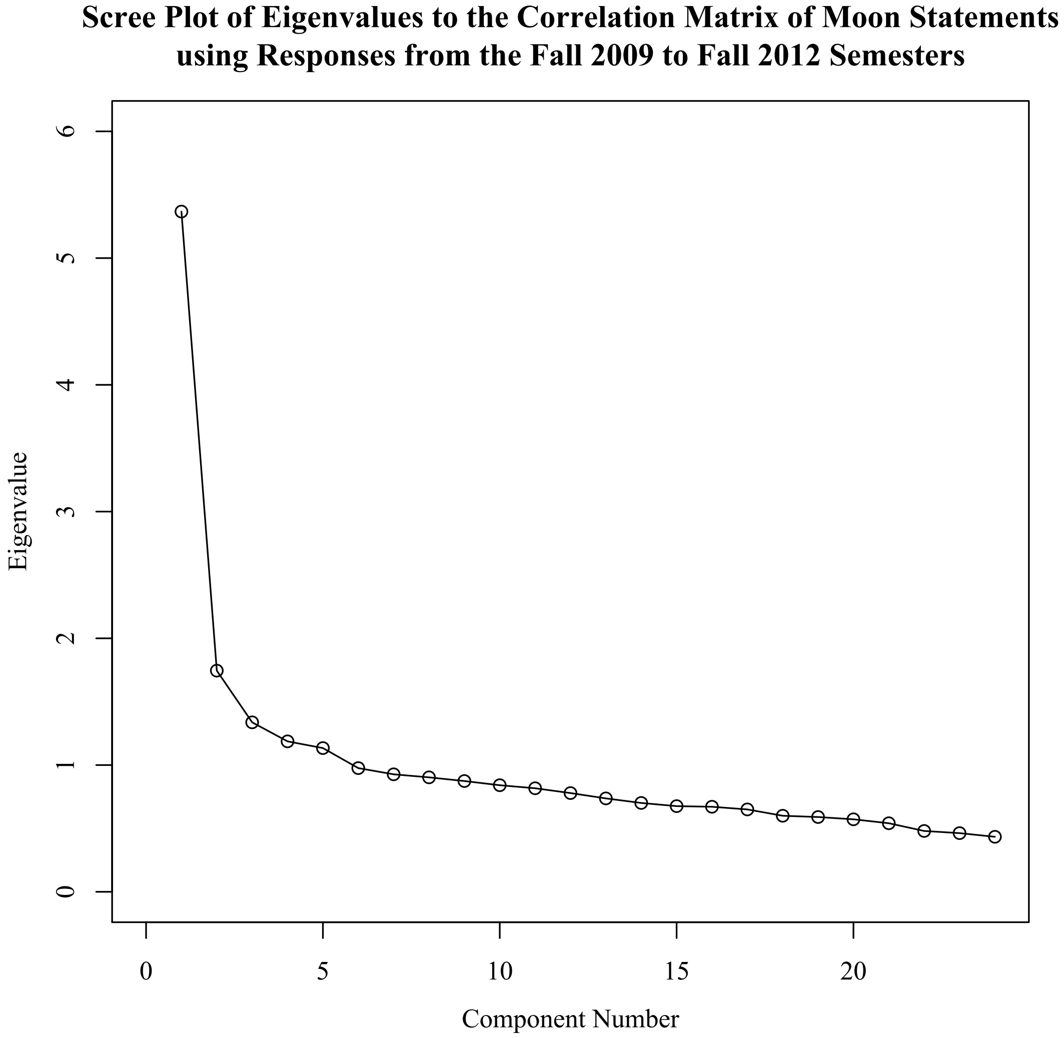


Figure 10.3. Scree plot of eigenvalues from the correlation matrix of responses to the statements about the Moon

I		II		III		IV		V	
sA91	**0.643**	**sA104**	**0.625**	**sA78**	**0.689**	**sA94**	**0.704**	**sA97**	**0.785**
sA89	**0.603**	**sA83**	**0.611**	**sA92**	**0.593**	**sA102**	**0.583**	**sA98**	**0.702**
sA90	**0.588**	**sA103**	**0.592**	**sA79**	**0.539**	**sA77**	**0.541**	sA88	0.422
sA87	**0.587**	**sA99**	**0.568**	**sA84**	**0.464**	**sA93**	**0.381**		
sA88	**0.510**	**sA100**	**0.452**	sA83	0.326	**sA96**	**0.341**		
sA80	**0.489**	**sA85**	**0.449**	sA77	0.315	sA103	0.402		
sA105	**0.471**	sA105	0.389	sA93	0.308	sA87	0.329		
		sA79	0.323			sA84	0.306		

Table 10.3. Reduced representation of factor loadings of statements about the Moon

In order of decreasing loading, the group of statements in the first factor consists of sA91, “the Moon has an atmosphere like the Earth,” sA89, “the Moon is about the same temperature as the Earth,” sA90, “the Moon has a helium atmosphere,” sA87, “the Moon has seas and oceans of water,” sA88, “the Moon is older than the Earth: a dead planet that used to be like Earth,” sA80, “craters are volcanic in origin,” and sA105, “the moon is lit by reflected ‘Earth light’ (that is, sunlight scattered off the Earth toward the Moon).” All statements in the first group except for sA80 pertain to comparisons of the Moon with the Earth, so the conceptual theme of the first group is essentially about surface feature comparisons between the Moon and the Earth.

The statements in the second group are sA104, “the side of the moon we don’t see is forever ‘dark’,” sA83, “the Moon is at a fixed distance from Earth,” sA103, “the Moon is larger at the horizon than when it is overhead,” sA99, “a lunar month is exactly 28 days long,” sA100, “at new Moon we are seeing the ‘far side’ of the Moon,” and sA85, “the Moon doesn’t rotate since we see only one side of it.” These statements essentially consist of statements pertaining to the phases of the Moon. Because these statements are grouped together, the most logical way to unlearn misconceptions about the phases of the Moon may be to devote a lecture to the phases of the Moon in which these misconceptions are addressed.

A brief thematic interpretation of the remaining groups follow. The statements in the third group are sA78, “the Moon doesn’t cause part of the tides,” sA92, “the

Moon has a smooth surface," sA79, "we see all sides of the Moon each month," and sA84, "the Moon changes physical shape throughout its cycle of phases." These statements are associated with the misconception about the physical shape of the Moon. The statements in the fourth group are sA94, "there is a real man in the Moon," sA102, "the Moon follows you in your car," sA77, "there is only one moon — ours," sA93, "the Moon sets during daylight hours and is not visible then," and sA96, "because the Moon reflects sunlight, it has a mirror-like surface." These statements are associated with misconceptions about the appearance of the full Moon from the Earth, that is, when the Moon is in its full Moon phase. The statements in the fifth group are sA97, "the Moon will someday crash into Earth," and sA98, "the Moon is a captured asteroid." These statements are associated with misconceptions about the origin and fate of the Moon.

10.5 Earth

I performed PCA on the ABI data for the 37 statements about the Earth from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 37 statements about the Earth. I determined the mean misconception retainment score for the four semesters ($M = 1.86$, $SD = 0.30$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 392. The internal consistency of the ABI responses was $\alpha = 0.89$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.4 shows the distribution of eigenvalues for statements about the Earth. Using the eigenvalue criterion, PCA supports nine groups of statements about the Earth from the extraction of nine factors, which cumulatively account for 53% of the total variance. The first factor accounts for 21% of the total variance, the second factor accounts for 8.3% of the variance, and the remaining factors account each account for up to 3.9% of the total variance. At the 50th percentile, extraction of the

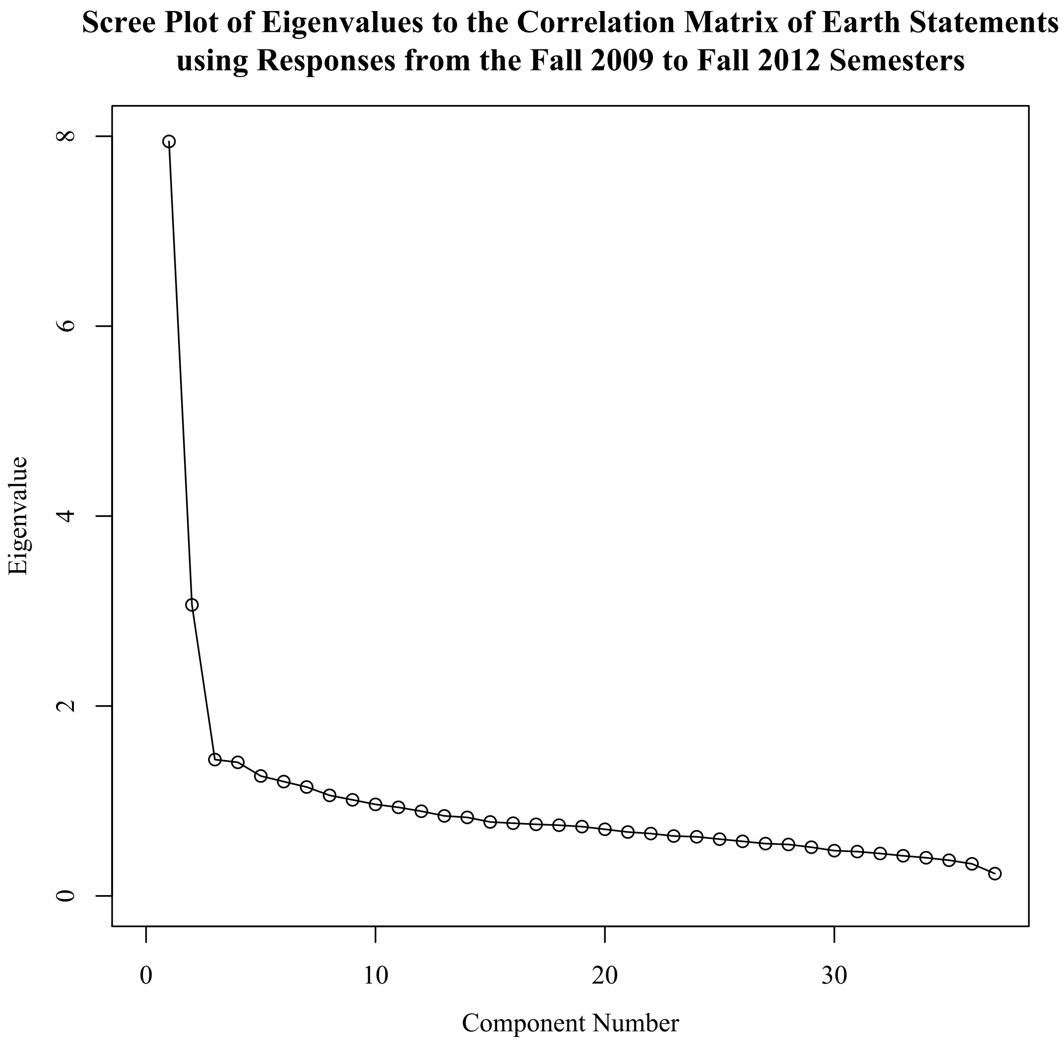


Figure 10.4. Scree plot of eigenvalues from the correlation matrix of responses to statements about the Earth

first two factors is consistent with the results of performing PA on the ABI data. Table 10.4 presents a reduced representation of factor loadings for statements about the Earth after extraction of nine factors.

I		II		III		IV		V	
sA156	**0.576**	**sA115**	**0.827**	**sA135**	**0.583**	**sA146**	**0.839**	**sA157**	**0.612**
sA159	**0.573**	**sA116**	**0.754**	**sA113**	**0.577**	**sA147**	**0.802**	**sA142**	**0.531**
sA137	**0.565**	**sA133**	**0.702**	**sA154**	**0.546**	**sA129**	**0.489**	**sA148**	**0.529**
sA152	**0.489**	**sA145**	**0.421**	**sA126**	**0.513**	sA144	0.360	**sA160**	**0.523**
sA143	**0.472**	sA160	0.504	**sA114**	**0.408**	sA122	0.331	**sA141**	**0.445**
sA128	**0.447**			sA143	0.374			sA128	0.374
sA122	**0.404**			sA129	0.330			sA131	0.311
sA158	**0.400**			sA142	0.324				
sA130	0.326								
sA141	0.319								
sA151	0.301								

VI		VII		VIII		IX	
sA118	**0.727**	**sA125**	**0.663**	**sA153**	**0.704**	**sA111**	**0.622**
sA149	**0.622**	**sA151**	**0.580**	**sA131**	**0.527**	**sA112**	**0.531**
sA144	**0.425**	**sA127**	**0.335**	**sA130**	**0.480**	sA114	0.371
sA150	0.333	sA126	0.410	**sA150**	**0.406**	sA158	0.340
sA158	0.319	sA114	0.370			sA148	0.334
		sA152	0.369			sA145	–0.306

Table 10.4. Reduced representation of factor loadings of statements about the Earth

In order of decreasing loading, the group of statements in the first factor consists of sA156, “the tides are caused just by the Earth’s rotation,” sA159, “tides are caused just by ocean winds,” sA137, “only Earth among the planets and moons has gravity,” sA152, “the Earth is the only planet with an atmosphere,” sA143, “the Earth will last forever,” sA128, “the Moon is not involved with any eclipses,” sA122, “X-rays can reach the ground,” and sA158, “the Sun is directly overhead everywhere on Earth at noon.” The first two statements in the group pertain to the tides, while the rest of the statements in the group pertain to a mixture of themes about the Earth’s atmosphere, the Earth’s gravity, the location of the Sun at noon, and whether or not the Earth will last forever. The underlying theme seems to connect with how the Earth is represented as if one views the Earth from outer space, e.g., with an atmosphere, and the Sun lighting part of the Earth. That two statements about the tides are included supports the notion that the Earth is viewed as a celestial object out there in space, because the Earth is often depicted with its tides drawn as bulging oceans on opposite sides of the Earth.

The statements in the second group are sA115, “Earth is at the center of the universe,” sA116, “Earth is the biggest planet,” sA133, “the sun orbits the Earth,” and sA145, “planes can fly in space.” These statements model the Earth as if it one stood on the Earth and observed everything else in space going around it. Interestingly, sA160, “the Earth is flat,” has a very significant loading on the second factor; however, sA160 loads the highest on the fifth factor and so is not, strictly speaking, part of the second group. Nonetheless, sA160 supports observations that the Earth appears flat when viewing from the ground, and that from the ground, planes appear to travel through the space above the Earth. Conceptually, the second group represents a set of statements pertaining to Earth-centered observations.

A brief thematic interpretation of the remaining groups follow. The statements in the third group are sA135, “solar eclipses happen about once a century and are seen everywhere on Earth,” sA113, “once ozone is gone from the Earth’s atmosphere, it will not be replaced,” sA154, “the Earth is not changing internally,” sA126, “you can see a solar eclipse from anywhere on Earth that happens to be facing the Sun at that time,” and sA114, “Earth and Venus have similar atmospheres.” There does not seem to be any particular theme for the statements in this group; however, these statements may be partly associated with misconceptions about solar eclipses and the composition of the Earth’s atmosphere and interior. The statements in the fourth group are sA146, “a day is exactly 24 hours long,” sA147, “a year is exactly 365 days long,” and sA129, “the day has always been 24 hours long.” These statements are associated with misconceptions about the length of a day and a year. The statements in the fifth group are sA157, “Earth has a second moon that only comes around once in awhile — ‘once in a blue moon’,” sA142, “meteorites have stopped falling onto the Earth,” sA148, “seasons are caused by speeding up and slowing down of Earth’s rotation,” sA160, “the Earth is flat,” and sA141, “seasons were chosen haphazardly.” There does not seem to be any particular theme for the statements in this group; however, these statements may be partly associated with misconceptions about the causes of seasons.

A thematic interpretation of the remaining groups continues. The statements in the sixth group are sA118, “Spring Tide is in the spring,” sA149, “the Earth orbits the sun at a constant speed,” and sA144, “the Earth’s magnetic poles go through its rotation poles.” These statements may be partly associated with misconceptions about assumptions made about the Earth based on the name of the terms given to it (e.g., “poles,” “spring tide”). The statements in the seventh group are sA125, “meteoroids enter the atmosphere a few times a night,” sA151, “the sky is blue because it reflects sunlight off oceans and lakes,” and sA127, “auroras are caused by sunlight reflecting off polar caps.” These statements are associated with misconceptions about the appearance of the sky. The statements in the eighth group are sA153, “comets affect the weather,” sA131, “Halley’s comet will eventually hit Earth,” sA130, “the air is a blue gas,” and sA150, “the Earth is in the middle of the Milky Way galaxy.” The first two of these statements may be partly associated with misconceptions about the influences of comets on the Earth. There does not otherwise appear to be any particular thematic interpretation for why these statements are grouped together. The statements in the ninth group are sA111, “Earth’s axis is not tilted compared to the ecliptic,” and sA112, “summer is warmer because we are closer to the sun during the summertime.” These statements are associated with misconceptions about the tilt of the Earth.

10.6 Sun

I performed PCA on the ABI data for the 32 statements about the Sun from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 32 statements about the Sun. I determined the mean misconception retainment score for the four semesters ($M = 1.91$, $SD = 0.29$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 396. The internal consistency of the ABI responses was $\alpha = 0.88$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.5 shows the distribution of eigenvalues for statements about the Sun. Using the eigenvalue criterion, PCA supports eight groups of statements about the Sun from the extraction of eight factors, which cumulatively account for 52% of the total variance. The first factor accounts for 23% of the total variance, while the remaining factors account each account for up to 6.8% of the total variance. At the 50th percentile, extraction of the first three factors is consistent with the results of performing PA on the ABI data. Table 10.5 presents a reduced representation of factor loadings for the sun statements after extraction of eight factors.

In order of decreasing loading, the group of statements in the first factor consists of sA189, “the Sun is the brightest star in universe,” sA190, “the Sun is the brightest object in the universe,” sA198, “the Sun is the largest star,” sA209, “the Sun is hottest star,” sA180, “the Sun is the hottest thing in the galaxy,” and sA177, “the Sun is a specific type of astronomical body with its own properties. It is not a star.” Interestingly, statement sA178, “the Sun will burn forever,” loads more strongly than sA177 on the first factor. The loading of sA178 on the second factor is only marginally higher than its loading on the first factor. Hence, PCA places sA178 in the second group of statements, although one may reasonably assert that sA178 has such a significantly high loading on the first factor that one may treat sA178 as if it is part of

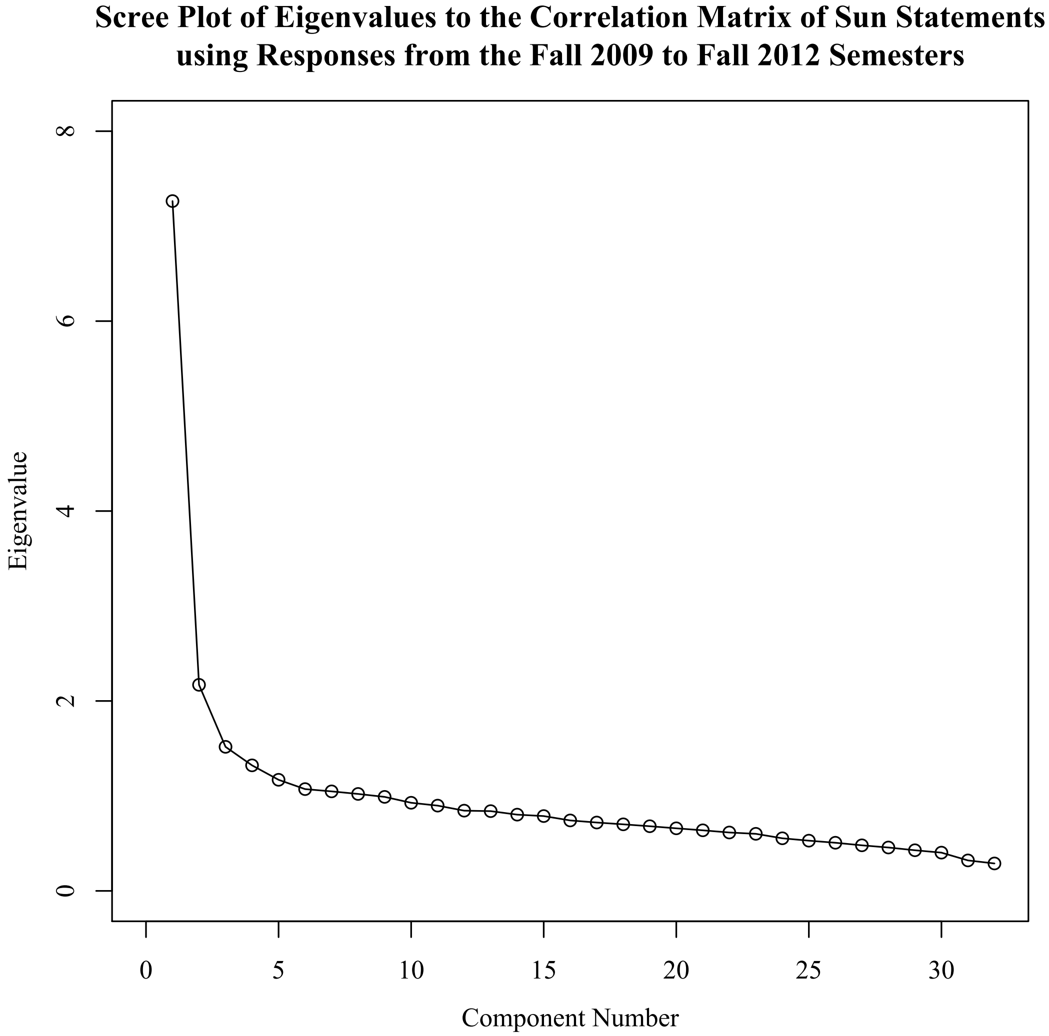


Figure 10.5. Scree plot of eigenvalues from the correlation matrix of responses to statements about the Sun

I		II		III		IV	
sA189	**0.774**	**sA192**	**0.686**	**sA199**	**0.611**	**sA193**	**0.666**
sA190	**0.748**	**sA206**	**0.556**	**sA200**	**0.553**	**sA196**	**0.599**
sA198	**0.742**	**sA178**	**0.433**	**sA187**	**0.537**	**sA214**	**0.549**
sA209	**0.701**	**sA181**	**0.418**	**sA204**	**0.438**	**sA215**	**0.399**
sA180	**0.679**	**sA182**	**0.413**	sA177	0.331	**sA211**	**0.385**
sA177	**0.373**	sA185	–0.374	sA183	0.329	sA177	0.317
sA178	0.428	sA188	0.321	sA186	0.303		

V		VI		VII		VIII	
sA184	**0.684**	**sA201**	**0.702**	**sA202**	**0.710**	**sA188**	**0.658**
sA191	**0.546**	**sA208**	**0.512**	sA217	0.412	**sA186**	**0.486**
sA185	**0.479**	**sA197**	**0.505**	sA215	0.398	sA177	–0.302
sA183	**0.479**	**sA217**	**0.427**	sA204	0.350		
sA213	**0.364**	sA187	0.321	sA183	0.318		
sA217	0.325			sA211	0.300		
sA181	0.323						

Table 10.5. Reduced representation of factor loadings of statements about the Sun

the first group. Comprehensively, the first group consists of statements where the Sun is modeled as the brightest, largest, and hottest of all the stars.

The statements in the second group are sA192, “the Sun is made of fire,” sA206, “the entire Sun is molten lava,” sA178, “the Sun will burn forever,” sA181, “the Sun does not move through space,” and sA182, “the Sun does not cause part of the tides.” The first three misconceptions model the Sun as a star that burns gas like fire. That sA181 is included suggests that students draw the associated misconceptions having already asserted that the Sun does not move. (Analogously, a fireplace does not move through one’s living room.) The second group consists of statements that depict the Sun as being on fire or that its light comes from burning its gas. That sA182 loads significantly on the second factor also suggests that students who endorse misconceptions about the Sun burning gas like fire also deny that the Sun contributes to tidal effects on the Earth. Interestingly, statement sA185, “the Sunspot cycle is 11 years long,” has a significant *negative* loading on the second factor. The significant negative loading suggests that students who reject the notion that the Sun burns gas like fire are actually *more* likely to endorse the misconception that the sunspot cycle is 11 years long.

The statements in the third group are sA199, “the Sun is hottest on its surface,” sA200, “the Sun has a solid core,” sA187, “Sunspots are constant fixtures on the sun,” and sA204, “the Sun’s surface is perfectly uniform.” These statements contrast physical features about the Sun’s surface vs. its core. Interestingly, students who endorse the misconception that the Sun’s core is solid are likely to endorse misconceptions about the Sun’s surface. Hence, the third group consists of statements regarding the Sun’s surface textures.

Notice that for the seventh factor, sA202, “the Sun is mostly iron,” is in a group by itself. Because PCA determines factor loadings for each statement, statement loadings are first determined, then the statements with the highest loadings on a given factor are grouped together. It turns out that sA202 does inter-correlate with other statements (e.g., sA204); however, those statements load more strongly on other factors. As a result, sA202 is in a group by itself.

A brief thematic interpretation of the remaining groups follow. The statements in the fourth group are sA193, “the Sun is a ‘heat planet’,” sA196, “the Sun is the smallest star in universe,” sA214, “the Sun is the only source of light in the galaxy — Sunlight reflects off planets and stars so we can see them,” sA215, “Sunspots are where meteors crash into the Sun,” and sA211, “it is possible that the Sun could explode in the ‘near future’.” There does not seem to be any particular theme for the statements in this group; however, these statements may be associated with uncertainty in how to characterize of the Sun, e.g., as a heat planet, as the source of light for other celestial bodies, or as something that might explode soon. The statements in the fifth group are sA184, “the Sun will blow up, become a black hole, and swallow the earth,” sA191, “the Sun always sets due west,” sA185, “the Sunspot cycle is 11 years long,” sA183, “sunspots are hot spots on the Sun’s surface,” and sA213, “the Sun doesn’t rotate.” Except for sA183, these statements are associated with misconceptions about cycles (e.g., sunspot cycle, day/night cycle, the Sun’s rotation) or the Sun as a threat to the Earth. The statements in the sixth group are sA201, “the Sun has only a few percent of the mass in the solar system,” sA208, “the Sun will explode as a nova,” sA197, “the Sun has no atmosphere,” and sA217, “it is more dangerous to look at the Sun during an eclipse because the radiation level from sun is greater then, than when there is no eclipse.” There does not seem to be any particular theme for the statements in this group. The statements in the eighth group are sA188, “the Sun is yellow,” and sA186, “the Sun’s surface temperature is millions of degrees Fahrenheit.” These statements are associated with misconceptions about the Sun’s surface color and temperature.

10.7 General astrophysics

I performed PCA on the ABI data for the 16 statements about astrophysics in general from the Fall 2009 to Fall 2012 semesters. Appendix A lists the 16 statements, under “General Astrophysics.” I determined the mean misconception retainment score for the four semesters ($M = 2.03$, $SD = 0.34$). For the purpose of performing PCA, missing cases were excluded listwise. The remaining sample size is 393. The internal consistency of the ABI responses was $\alpha = 0.80$, which, as discussed in Section 3.4, is considered adequate.

Figure 10.6 shows the distribution of eigenvalues for statements about astrophysics in general. Using the eigenvalue criterion, PCA supports four groups from the extraction of four factors, which cumulatively account for 47% of the total variance. The first factor accounts for 25% of the total variance, while the remaining factors account each account for up to 8.1% of the total variance. At the 50th percentile, extraction of the first two factors is consistent with the results of performing

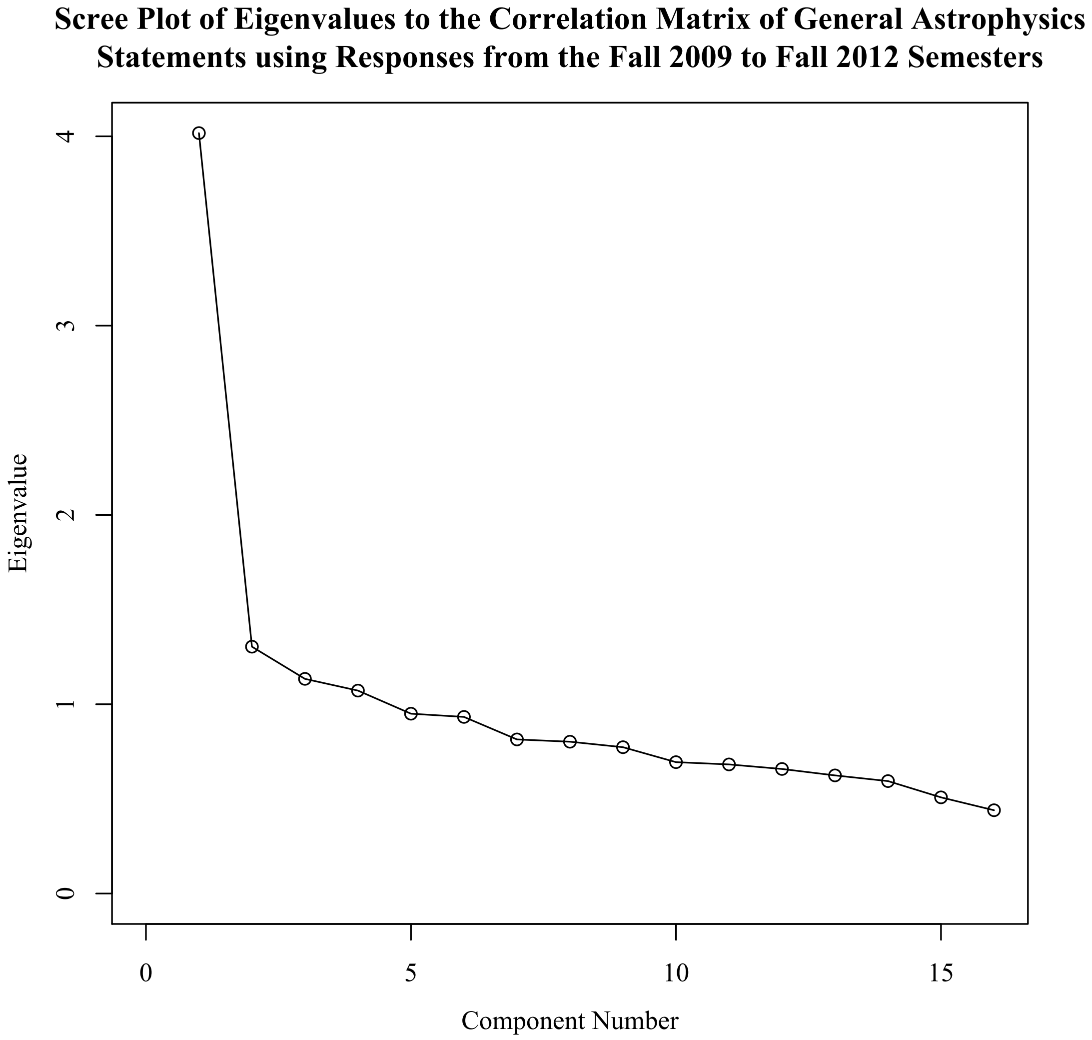


Figure 10.6. Scree plot of eigenvalues from the correlation matrix of responses to statements about astrophysics in general

PA on the ABI data. Table 10.6 presents a reduced representation of factor loadings for the General Astrophysics statements after extraction of four factors.

I		II		III		IV	
sA256	**0.702**	**sA262**	**0.688**	**sA271**	**0.685**	**sA248**	**0.652**
sA267	**0.576**	**sA254**	**0.603**	**sA273**	**0.659**	**sA253**	**0.636**
sA255	**0.535**	**sA270**	**0.595**	**sA272**	**0.513**	**sA261**	**0.488**
sA263	**0.512**	**sA258**	**0.461**	**sA252**	**0.493**	sA258	0.378
sA259	**0.512**	sA255	0.426	sA259	0.305	sA254	0.371
sA261	0.396	sA252	0.382			sA272	0.347

Table 10.6. Reduced representation of factor loadings of statements about astrophysics in general

In order of decreasing loading, the group of statements in the first factor consists of sA256, “space is infinite,” sA267, “there is a center to the universe,” sA255, “light travels infinitely fast,” sA263, “astronomical ideas of mass, distance, and temperature of planets are all speculative,” and sA259, “gravity is the strongest force in the universe.” These statements model the fundamental properties of the universe as a whole. Interestingly, students who endorse this incorrect interpretation of our universe also think that mass, distance, and temperature of planets are just speculative quantities, rather than fundamental quantities.

The statements in the second group are sA262, “the universe as a whole is static (unchanging),” sA254, “satellites need continuous rocket power to stay in orbit around the Earth,” sA270, “smaller telescopes enable astronomers to see smaller details,” and sA258, “telescopes cannot see any details on any of the planets.” These statements pertain to our ability to document the universe, essentially with telescopes in orbit around the Earth. To a lesser extent, statements sA255, “light travels infinitely fast,” and sA252, “astronomy and astrology are the same thing,” load on the second factor. Statement sA255 may naturally be expected to load with statements about telescopic observations of the universe. That sA252 loads significantly on the second factor, however, also implies that students who endorse misconceptions about documenting the universe with space telescopes also tend to see astronomy and astrology as the same subject.

The statements in the third group are sA271, “the most important function of a telescope is magnification,” sA273, “astronomers mostly work with telescopes,” sA272, “all space debris existing today is the result of planet collisions and explosions on planets,” and sA252, “astronomy and astrology are the same thing.” The two highest-loading statements, sA271 and sA273, refer to the utility of telescopes. Statement sA252 loads the highest on the third factor, which suggests that students who misunderstand the utility of telescopes also tend to see astronomy and astrology as the same science. Strangely, statement sA272, which loads the highest on the third factor, has nothing to do with telescopes, but instead pertains to the origin of space debris. The third factor essentially consists of statements about the purpose of a telescope in studying astronomy. Interestingly, PCA *does not* combine the telescope statements from the second and third factors, which suggests that, in the minds of the students, the process of scientific documentation of the universe using space telescopes is conceptually different from the role and function of telescopes in general.

The statements in the fourth group are sA248, “cosmic rays are light rays,” sA253, “gravity will eventually pull all the planets together,” and sA261, “we can hear sound in space.” To a lesser extent, sA254, sA258, and sA272 load on the fourth factor. These statements pertain to misconceptions about particular celestial objects or particles in the universe.

CHAPTER 11

PRINCIPAL COMPONENTS ANALYSIS ON COMBINED INVENTORY TOPICS

11.1 Overview

In Chapters 8-10, I used PCA to produce factor loadings for each topic of the ABI. The resulting groups suggest the most logical subtopics to which one should teach, since the unlearning of the misconceptions associated with each subtopic item are the most inter-correlated. Each topic of the ABI, however, was analyzed one at a time. In this Chapter, I examine the extent to which misconceptions are embedded in themes represented by various statements throughout multiple topics of the ABI. As with my analysis in Chapters 8-10, I will continue to use data from the Fall 2009 to Fall 2012 semesters. My analysis is broken up into the following combined topics:

1. Stars and Sun
2. Stars and Galaxies
3. Stars and Black Holes
4. Solar System and Planets
5. Planets and Earth
6. Moon and Earth
7. Earth and Sun
8. Moon Earth Sun
9. Galaxies and Black Holes
10. Galaxies and Cosmology

The motivation behind the analysis is as follows: PCA can predict whether or not groups of misconceptions are highly inter-correlated between or among topics. If, after performing PCA, for example, misconceptions about stars (other than the Sun), from the Stars topic of the ABI are highly inter-correlated with misconceptions about the Sun, then one should teach to the relevant misconceptions by making targeted associations between the Sun and other stars. On the other hand, if after performing PCA, misconceptions about the stars and misconceptions about the Sun appear in separate groups within the factor structure, then there is little connection between misconceptions about stars in general and misconceptions about the Sun. In such a case, the stars and the Sun should be treated as separate topics during instruction. Likewise, PCA can make this distinction for other combined topics of the ABI.

In my subsequent analysis, for each of these topics, I indicate the number of factors that one would extract under the eigenvalue criterion. Within those factors, I present and discuss the results exclusively for a subset of the groups of misconceptions that arise from the factor structure. I chose to discuss a subset of the factors, for two reasons. First, the statistical power (the sample size to the number of statements) from combining topics in the ABI is reduced because of the increase in the number of statements in the analysis. As discussed in Chapter 10, analysis of the data from a single topic of the ABI may have a statistical power as low as 11. Second, a comprehensive discussion of all of the factors is not particularly insightful. As discussed in Section 6.3.3, the methodology of parallel analysis (PA) is that only a subset of all of the factors that one could extract are particularly meaningful. These factors include groups of misconceptions that are unlikely to arise by random sampling of the data. Therefore, I restrict my subsequent analysis of the groups of misconceptions to just those within the number of factors supported by PA.

In the analysis that follows, I present the number of factors to extract using the criteria by Kaiser and the PA technique, and I present the groups of misconceptions for each of the factors supported by PA. Within the groups, factor loadings are included but not discussed, since I intend simply to argue whether or not various topics should be mixed. A symbol to identify the topic from which each statement in the group was taken is presented with each of the statements. The list of symbols is presented in Table 11.1. The Planets topic consists of the aggregate of all statements from the Venus, Mars, and Saturn topics (but not the Earth topic). The Cosmology topic consists of five statements about cosmology from the topic labeled General Astrophysics: sA253, sA256, sA262, sA263, and sA267. The reader may refer to Appendix A for a list of statements in the ABI.

Topic	**Symbol**
Stars	★
Solar System	§
Moon	◑
Venus	♀
Earth	⊕
Mars	♂
Saturn	♄
Planets	P
Sun	⊙
Galaxies	G
Black Holes	●
Cosmology	C

Table 11.1. Symbols for the topics of the ABI

11.2 The Sun and other stars

I performed PCA on the aggregate of statements about the Sun and other stars, respectively with 32 and 36 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 19 groups from the extraction of 19 factors, of which the first five are supported by PA. The first factor accounts for 18% of the total variance, and the first five factors cumulatively account for 32% of the total variance. Table 11.2 presents the statements in the first five groups.

I			II			III			IV		
⊙	sA189	0.774	★	sA34	0.716	★	sA14	0.686	★	sA35	0.653
⊙	sA198	0.744	★	sA22	0.597	★	sA3	0.641	★	sA28	0.647
⊙	sA190	0.728	★	sA33	0.593	★	sA15	0.612	★	sA21	0.529
⊙	sA209	0.696	★	sA20	0.585	★	sA4	0.476	★	sA30	0.416
⊙	sA180	0.611	★	sA23	0.473	★	sA11	0.399	★	sA37	0.405
★	sA17	0.539				★	sA18	0.366	★	sA29	0.369
★	sA6	0.392				★	sA24	0.300			
⊙	sA181	0.298									

V		
⊙	sA217	0.679
⊙	sA208	0.583
⊙	sA211	0.405

Table 11.2. Reduced representation of statement factor loadings regarding the Sun and other stars

Table 11.2 shows that for the most part, misconceptions about stars are to be treated almost exclusively separate from misconceptions about the Sun. The second through fourth groups consist of only misconceptions about stars, while misconceptions about the Sun appear primarily in the fifth through eighth groups (of which only the fifth group is shown).

The first factor also consists primarily of Sun statements, with weaker loadings for the star statements. The statements about the Sun in this factor are sA189, “the Sun is the brightest star in universe,” sA198, “the Sun is the largest star,” sA190, “the Sun is the brightest object in the universe,” sA209, “the Sun is hottest star,” sA180, “the Sun is the hottest thing in the galaxy,” and, to a much lower extent, sA181, “the Sun does not move through space.” The first five of these statements about the Sun are also the same first five statements that appear in the first group of Sun statements in my PCA of the Sun topic, as discussed in Section 10.6. The two star statements in the first group are sA17, “all stars are smaller than the Sun,” and sA6, “we are looking at stars as they are now.” Essentially, the first factor of statements about the stars and the Sun can be thought of as treating the Sun as being brighter, larger, and hotter than all other stars. In essence, PCA says that misconceptions about the stars and the Sun are best unlearned when the two topics are treated separately, with the exception of comparing the size, temperature, and brightness of the Sun to other stars in the universe.

One might consider the possibility that the groups in Table 11.2 may have arisen largely because the items were originally presented as statements within topics, in the original form of the ABI. As I show in subsequent sections, however, when pooling together enough statements from multiple topics, particular themes emerge in which statements from more than one topic are grouped together. For example, in Section 11.6, when statements about the Earth are included with statements about Venus, Mars, and Saturn, the four factors which account for the highest variance have groups that contain statements about the Earth and at least one of the other planets under consideration. As another example, in Section 12.2, I introduce a method that reduces the original set of 215 statements to a set of 27 statements, while still preserving common themes of misconceptions. The process produces four groups of inter-correlation statements from the set of 27 statements, and each group contains statements from at least two topics. Hence, *the groups that result from performing PCA on aggregates of topics may arise particularly because of the presence of particular themes of misconceptions endorsed by the students.*

11.3 Stars and galaxies

I performed PCA on the aggregate of statements about stars and galaxies, respectively with 34 and 12 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 12 groups from the extraction of 12 factors, of which the first three are supported by PA. The first factor accounts for 21% of the total variance, and the first three factors cumulatively account for 29% of the total variance. Table 11.3 presents the statements in the first three groups.

I			II			III		
★	sA11	0.613	★	sA34	0.687	G	sA228	0.635
★	sA17	0.607	★	sA33	0.635	G	sA230	0.616
★	sA18	0.504	★	sA22	0.487	G	sA227	0.570
★	sA15	0.493	G	sA231	0.402	G	sA225	0.542
★	sA20	0.458	★	sA31	0.367			
★	sA7	0.366						
★	sA6	0.363						

Table 11.3. Reduced representation of statement factor loadings regarding stars and galaxies

Table 11.3 shows that misconceptions about stars are to be treated almost exclusively separate from misconceptions about galaxies. The first two groups of statements pertain to misconceptions about stars. The first group of statements consists of sA11, “stars last forever,” sA17, “all stars are smaller than the Sun,” sA18, “the galaxy, solar system and universe are the same things,” sA15, “all stars have same color and size,” sA20, “stars just existed — they don’t make energy or change size or color,” sA7, “stars actually twinkle — change brightness,” and sA6, “we are looking at stars as they are now.” Five of these statements comprise the second group of star statements in my analysis of exclusively the Stars topic, in Section 10.2. The third group consists primarily of three galaxy statements pertaining to visual properties of galaxies, as discussed in Section 9.1.

11.4 Stars and black holes

I performed PCA on the aggregate of statements about stars and black holes, respectively with 34 and 13 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 14 groups from the extraction of 14 factors, of which the first four are supported by PA. The first factor accounts for 19% of the total variance, and the first four factors cumulatively account for 32% of the total variance. Table 11.4 presents the statements in the first four groups.

I			II			III			IV		
★	sA34	0.677	★	sA14	0.670	●	sA233	0.689	★	sA28	0.597
★	sA20	0.622	★	sA3	0.652	●	sA246	0.591	★	sA35	0.597
★	sA22	0.618	★	sA15	0.632	●	sA232	0.554	★	sA21	0.528
★	sA23	0.559	★	sA18	0.500	●	sA247	0.507	★	sA30	0.493
★	sA33	0.456	★	sA17	0.469	●	sA245	0.437	★	sA24	0.455
			★	sA4	0.390				★	sA37	0.426
			★	sA11	0.379				★	sA29	0.412

Table 11.4. Reduced representation of statement factor loadings regarding stars and black holes

Table 11.4 shows that misconceptions about stars are to be treated exclusively separate from misconceptions about black holes. The first, second, and fourth groups consist of exclusively star statements, and the third group consists of exclusively black hole statements, four of which appear in the group of black hole statements labeled "Appearance and Fate" in my PCA of the Black Holes topic, as discussed in Section 9.2.

11.5 Solar system and planets

I performed PCA on the aggregate of statements about solar system and planets, respectively with 30 and 17 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 14 groups from the extraction of 14 factors, of which the first five are supported by PA. The first factor accounts for 18% of the total variance, and the first five factors cumulatively account for 34% of the total variance. Table 11.5 presents the statements in the first five groups.

Table 11.5 shows that misconceptions about the solar system are to be treated almost exclusively separate from misconceptions about the planets. The first group of statements consists exclusively of misconceptions about Mars, the second group of statements consists almost exclusively of same three misconceptions about Saturn as indicated in the second group in Table 8.3, and the third and fifth groups of statements consist almost exclusively of misconceptions about the solar system. The fourth group

I			II			III			IV		
♂	sA167	0.796	♄	sA172	0.710	§	sA60	0.601	♀	sA106	0.706
♂	sA168	0.758	♄	sA171	0.702	§	sA59	0.544	♀	sA109	0.572
♂	sA161	0.537	♄	sA176	0.633	§	sA62	0.526	♀	sA107	0.567
♂	sA169	0.488	§	sA50	0.339	§	sA56	0.501	♂	sA165	0.401
♂	sA164	0.422				§	sA63	0.447			
						§	sA41	0.406			
						§	sA57	0.388			

V		
§	sA51	0.641
§	sA53	0.634
§	sA43	0.434
♂	sA166	0.385

Table 11.5. Reduced representation of statement factor loadings regarding the solar system and planets

consists of three statements about Venus and the weaker-loading statement about Mars, sA165, “Mars could be made inhabitable.”

In Chapter 8, I showed that 11 of 17 misconceptions about Venus, Mars, and Saturn are best unlearned by teaching the properties of the associated planets, one planet at a time, whereas the remaining six misconceptions are best unlearned by making targeted comparisons between their surface features and the relation of the planets to the Earth. The results presented in Table 11.5 support that planet misconceptions are primarily grouped together, once again by planet, and should not be mixed in with misconceptions about the solar system. Therefore, *my PCA results on the combined planets-solar system topics support that each planet should be taught one at a time*.

11.6 The Earth and other planets

I performed PCA on the aggregate of statements about the Earth and three other planets (Venus, Mars, and Saturn), respectively with 37 and 17 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 13 groups from the extraction of 13 factors, of which the first five are supported by PA. The first factor accounts for 21% of the total variance, and the first five factors cumulatively account for 37% of the total variance. Table 11.6 presents the statements in the first five groups.

I			II			III			IV		
⊕	sA115	0.788	♂	sA167	0.725	♀	sA109	0.704	♄	sA172	0.671
⊕	sA116	0.712	♂	sA168	0.692	⊕	sA114	0.669	♄	sA171	0.598
⊕	sA133	0.695	♂	sA161	0.684	♀	sA107	0.637	♄	sA176	0.588
⊕	sA160	0.609	⊕	sA157	0.608	⊕	sA126	0.394	⊕	sA143	0.497
♂	sA166	0.449	⊕	sA128	0.399	⊕	sA127	0.385	♂	sA169	0.402
⊕	sA145	0.439							⊕	sA152	0.380
⊕	sA148	0.365									

5		
⊕	sA118	0.703
⊕	sA149	0.533
⊕	sA144	0.498
⊕	sA135	0.458
⊕	sA113	0.404
⊕	sA154	0.379

Table 11.6. Reduced representation of statement factor loadings regarding the Earth and other planets

Table 11.6 shows that misconceptions about the Earth tend to mix with misconceptions about other planets, with only the fifth factor consisting of exclusively statements about the Earth. The first group consists of six Earth statements and one Mars statement. In the first group, the statements about the Earth are sA115, “Earth is at the center of the universe,” sA116, “Earth is the biggest planet,” sA133, “the sun orbits the Earth,” sA160, “the Earth is flat,” sA145, “planes can fly in space,” and sA148, “seasons are caused by speeding up and slowing down of Earth’s rotation.” The first five of these statements appear in the second group of Earth statements in my analysis of exclusively the Earth topic, as discussed in Section 10.5. The one statement about Mars is sA166, “Mars is the second largest planet.” Comprehensively, these statements treat the Earth as the planet around which objects in space are observed to move. PCA thus suggests that the most effective way to dispel the misconception about Mars being the second largest planet is to show, additionally, why Earth is not the biggest planet. For example, a drawn-to-scale comparison of the sizes of the Earth, Mars, and the outer planets may illustrate the relative sizes of the planets efficiently.

The second group consists of three Mars statements and two Earth statements. In the second group, the statements about Mars are sA167, “life, when it did exist on Mars, was quite advanced,” sA168, “there are Lowellian canals on Mars built by intelligent beings,” and sA161, “Mars is green (from plant life).” The two statements about the Earth are sA157, “Earth has a second moon that only comes around once in awhile — ‘once in a blue moon’,” and sA128, “the Moon is not involved with any eclipses.” The first three statements pertain to life on Mars, while the two Earth statements actually draw attention to the Moon, or a second moon around the Earth. There does not seem to be any obvious construct that represents all five of the statements in the group. Since sA128 has a substantially lower factor loading than the others in the group, the misconception associated with sA128 does not connect as strongly to the other misconceptions in the group. Essentially, PCA suggests that the misconception about a second moon around the Earth is best unlearned by also teaching in the context of misconceptions about life on Mars.

The third group consists of two Venus statements and three Earth statements. In the third group, the statements about Venus are sA109, “Venus is a lot like the Earth in temperature,” and sA107, “clouds on Venus are composed of water, like clouds on earth.” The statements about the Earth are sA114, “Earth and Venus have similar atmospheres,” sA126, “you can see a solar eclipse from anywhere on Earth that happens to be facing the Sun at that time,” and sA127, “auroras are caused by sunlight reflecting off polar caps.” Note that the factor loadings for sA126 and sA127 are substantially lower than those of sA114 and the two Venus statements in the group. Hence, the three most inter-correlated statements make comparisons about the atmospheres between Venus and the Earth, so the third factor consists of statements which compare the atmospheres of the Earth and Venus.

The fourth group consists of three statements about Saturn, two statements about the Earth, and one statement about Mars. In the fourth group, the statements about Saturn are sA172, “Saturn’s rings are solid,” sA171, “Saturn is the only planet with rings,” and sA176, “Saturn has only one ring.” The statements about the Earth are sA143, “the Earth will last forever,” and sA152, “the Earth is the only planet with an atmosphere.” The statement about Mars is sA169, “Mars is Hot because it is red ... Mars — god of fire.” The statements that load the highest pertain to Saturn’s rings. Interestingly, students who endorse misconceptions about Saturn’s rings also tend to assume that the Earth will last forever and, to a lesser extent, only the Earth has an atmosphere. Comprehensively, the fourth group of statements represents misconceptions about the rings of Saturn. To a lesser extent, misconceptions about Saturn’s rings are linked with the misconceptions that the Earth will last forever and is the only planet with an atmosphere, and that Mars is hot because of its red surface. No misconceptions about the atmosphere of Venus load significantly on the fourth factor.

In short, PCA suggests that *targeted comparisons between the surfaces of other worlds may help to dispel misconceptions that students have about some surface properties of Venus, the Earth, Mars, and Saturn.* The first group supports teaching about the scale sizes of the planets. The second group supports teaching in the context of misconceptions about life on Mars together with the misconception about a second Moon around the Earth. The third group supports comparisons between the atmospheres of Venus and the Earth. The fourth group supports teaching primarily about the rings and atmosphere of Saturn, with the option to show that Mars is not hot even though its surface is red.

11.7 The Moon and the Earth

I performed PCA on the aggregate of statements about the Moon and the Earth, respectively with 24 and 37 statements in the associated topics in the ABI. Using the eigenvalue criterion, PCA supports 17 groups from the extraction of 17 factors, of which the first three are supported by PA. The first factor accounts for 20% of the total variance, and the first three factors cumulatively account for 30% of the total variance. Table 11.7 presents the statements in the first three groups.

	I			II			III	
⊕	sA115	0.809	⊕	sA159	0.648	⊕	sA125	0.636
⊕	sA116	0.714	⊕	sA156	0.640	⊕	sA100	0.573
⊕	sA133	0.677	⊕	sA128	0.442	⊕	sA114	0.531
⊕	sA160	0.511	⊕	sA122	0.439	⊕	sA127	0.454
⊕	sA145	0.395	◑	sA92	0.402	⊕	sA126	0.413
◑	sA94	0.314	◑	sA96	0.338			

Table 11.7. Reduced representation of statement factor loadings regarding the Moon and the Earth

While not explicitly shown in Table 11.7, the fourth group consists of exclusively Moon statements: sA91, sA89, sA87, sA88, and sA80. Together with Table 11.7, misconceptions about the Earth are, for the most part, to be treated separately from misconceptions about the Moon. The first three groups consists of primarily Earth statements. Four of the statements in the first group consists of the same statements as those in the second group of Table 11.6, which treat the Earth as the center around which objects in space move. That nearly the same group of Earth statements appears in both tables supports that these misconceptions about the Earth can be untaught without making reference to misconceptions about other planets or the Moon.

The second group consists of four Earth statements and two Moon statements. In the second group, the statements about the Earth are sA159, “tides are caused just by ocean winds,” sA156, “the tides are caused just by the Earth’s rotation,” sA128, "the Moon is not involved with any eclipses," and sA122, "X-rays can reach the ground." These statements pertain primarily to things that could influence tides and eclipses on the Earth. The statements about the Moon are sA92, "the Moon has a smooth surface," and sA96, "because the Moon reflects sunlight, it has a mirror-like surface," These statements pertain to the surface texture of the Moon. Comprehensively, the statements in the second group seem to utilize knowledge about the lit and unlit sides of the Earth and the Moon. PCA thus suggests that misconceptions about the orientation and phases of the Earth and the Moon may be significantly inter-correlated, so one should teach to these misconceptions together.

11.8 The Earth and the Sun

I performed PCA on the aggregate of statements about the Earth and the Sun, respectively with 37 and 32 statements in the associated topics. Using the eigenvalue criterion, PCA supports 19 groups from the extraction of 19 factors, of which the first four are supported by PA. The first factor accounts for 20% of the total variance, and the first four factors cumulatively account for 32% of the total variance. Table 11.8 presents the statements in the first four groups.

	I			II			III			IV	
⊙	sA189	0.760	⊕	sA115	0.789	⊕	sA146	0.807	⊙	sA192	0.665
⊙	sA190	0.715	⊕	sA116	0.777	⊕	sA147	0.763	⊙	sA206	0.596
⊙	sA198	0.711	⊕	sA133	0.743	⊕	sA129	0.565	⊕	sA143	0.539
⊙	sA209	0.711	⊕	sA160	0.643	⊙	sA191	0.333	⊙	sA178	0.449
⊙	sA180	0.619	⊕	sA128	0.415				⊙	sA193	0.370
			⊕	sA148	0.409				⊙	sA181	0.362
			⊕	sA145	0.386				⊙	sA177	0.336
			⊕	sA147	0.345						

Table 11.8. Reduced representation of statement factor loadings regarding the Earth and the Sun

Table 11.8 shows that misconceptions about the Earth are to be treated almost exclusively separate from misconceptions about the Sun. The first group consists exclusively of those Sun statements which I also previously identified in my analysis of exclusively the Sun topic, in Section 10.6, which treat the Sun as the brightest, largest, and hottest of all the stars. In the second group, the four highest-loading statements consist of those Earth statements that deal with observing objects going around the Earth, as discussed in Section 11.6. The third group consists of three Earth statements which appear in the fourth group of Earth statements in Table 10.4. These statements are: sA146, “a day is exactly 24 hours long,” sA147, “a year is exactly 365 days long,” and sA129, “the day has always been 24 hours long.”

In Table 11.8, the fourth group consists of six Sun statements and one Earth statement. In the fourth group, the statements about the Sun are sA192, “the Sun is made of fire,” sA206, “the entire Sun is molten lava,” sA178, “the Sun will burn forever,” sA193, “the Sun is a ‘heat planet’,” sA181, “the Sun does not move through space,” and sA177, “the Sun is a specific type of astronomical body with its own properties. It is not a star.” Four of these statements appear in the second group of statements in my analysis of exclusively the Sun topic, as discussed in Section 10.6. The statement about the Earth is sA143, “the Earth will last forever.” PCA suggests that students who endorse misconceptions about the composition or movement of the Sun also tend to think that the Earth will last forever.

11.9 The Moon, the Earth, and the Sun

I performed PCA on the aggregate of statements about the Moon, the Earth, and the Sun, respectively with 24, 37, and 32 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports 27 groups from the extraction of 27 factors, of which the first four are supported by PA. The first factor accounts for 19% of the total variance, and the first four factors cumulatively account for 30% of the total variance. Table 11.9 presents the statements in the first four groups.

I			II			III			IV		
⊙	sA189	0.764	⊕	sA115	0.783	⊕	sA146	0.822	⊕	sA143	0.589
⊙	sA190	0.730	⊕	sA116	0.729	⊕	sA147	0.782	⊙	sA206	0.567
⊙	sA198	0.711	⊕	sA133	0.720	⊕	sA129	0.514	⊙	sA192	0.517
⊙	sA209	0.704	⊕	sA160	0.618	⊙	sA191	0.389	⊙	sA178	0.476
⊙	sA180	0.630	⊕	sA148	0.413	◑	sA83	0.303	⊙	sA204	0.295
⊙	sA177	0.311	⊕	sA145	0.362	⊕	sA144	0.296	◑	sA99	–0.294

Table 11.9. Reduced representation of statement factor loadings regarding the Moon, the Earth, and the Sun

Table 11.9 shows that misconceptions about the Moon, the Earth, and the Sun sometimes tend to mix. The first group consists of six Sun statements, five of which are the same as those in the first group of Table 11.8, which reinforces that misconceptions about the Sun as the brightest, largest, and hottest of all the stars should be targeted without referring to misconceptions about the Earth or the Moon. The second group consists of six Earth statements, all of which appear in the first group of Table 11.6 and the second group of Table 11.8, which reinforces that misconceptions which treat the Earth as the center around which objects move should be targeted without referring to misconceptions about the Moon or the Sun.

In Table 11.9, the third group consists of four Earth statements, one Sun statement, and one Moon statement. In the third group, the statements about the Earth are sA146, “a day is exactly 24 hours long,” sA147, “a year is exactly 365 days long,” sA129, “the day has always been 24 hours long,” and sA144, “the Earth’s magnetic poles go through its rotation poles.” The first three of these statements also appeared in the third group of mixed Earth and Sun statements, in Table 11.8. Hence PCA tends to preserve that the misconceptions associated with these statements remain correlated despite the aggregation of Earth, Moon, and Sun statements. In Table 11.9, the third group also consists of sA191, “the Sun always sets due west,” and sA83, “the Moon is at a fixed distance from Earth.” Comprehensively, the third group treats the Earth-Moon-Sun system as never changing.

The fourth group consists mostly of those four Sun statements that depict the Sun as being on fire or that its light comes from burning its gas. Statements that relate closely to this theme have been previously identified in the second group of Table 10.5 and the fourth group of Table 11.8. In Table 11.9, the fourth group consists of sA143, “the Earth will last forever,” and sA99, “a lunar month is exactly 28 days long,” which loads the least and has a negative factor loading. Comprehensively, the fourth group preserves the theme of treating the Sun as a forever-burning ball of gas.

11.10 Galaxies and black holes

I performed PCA on the aggregate of statements about galaxies and black holes, respectively with 12 and 13 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports five groups from the extraction of five factors, of which the first three are supported by PA. The first factor accounts for 26% of the total variance, and the first three factors cumulatively account for 38% of the total variance. Since there are only five factors in total, however, I will briefly discuss all five resulting statement groups in the subsequent discussion. Table 11.10 presents the statements in all five groups.

Table 11.10 shows that misconceptions about galaxies sometimes tend to mix with misconceptions about black holes. The first group of statements consists of six of the same galaxy statements which I identified previously in the factor under the label “Visual Properties” in my analysis of exclusively the Galaxies topic, in Table 9.4. Mixed in with these statements is sA234, “black holes really don’t exist,” which suggests that students who endorse misconceptions about the visual properties of galaxies also do not think that black holes really exist.

The second group of statements consists of two galaxy statements and three black hole statements. In the second group, the two galaxy statements are sA226, “the galaxies are randomly distributed,” and sA227, “we see all the stars that are in the

I			II			III			IV		
G	sA228	0.628	G	sA226	0.662	●	sA237	0.681	G	sA224	0.781
G	sA230	0.586	●	sA242	0.527	●	sA238	0.637	G	sA222	0.740
G	sA219	0.562	G	sA227	0.488	●	sA235	0.629	G	sA221	0.702
G	sA231	0.549	●	sA233	0.474	●	sA246	0.435			
G	sA225	0.524	●	sA232	0.457	G	sA220	0.396			
●	sA234	0.522									
G	sA218	0.417									

V		
●	sA244	0.765
●	sA240	0.736
●	sA243	0.463
●	sA247	0.455
●	sA245	0.432

Table 11.10. Reduced representation of statement factor loadings regarding galaxies and black holes

Milky Way," the same two statements that comprise the two-statement group in the Galaxies factor labeled "Spatial Distribution," in Table 9.4. The statements about black holes are sA242, "black holes can be seen visually, like seeing a star or planet," sA233, "black holes last forever," and sA232, "black holes create themselves from nothing," which were also identified in the group under the Black Holes factor label "Appearance and Fate," in Table 9.10. PCA thus suggests that misconceptions about galaxy spatial distribution and misconceptions about the appearance and fate of black holes may tend to reinforce each other. In other words, that a student endorses "galaxies are randomly distributed" also implies endorsement of misconceptions related to the appearance and fate of black holes. One might suppose that because we first learn about galaxies and black holes through some sort of visual media (e.g., as drawings in books, including the textbook used in a typical introduction to astronomy course, or as a slide on a projector), students may refer back to their own recollection of how galaxies and black holes were presented to them. Conceptually, the second group may consist of statements about how students recall the appearance of galaxies and black holes in visual media.

The third group of statements consists of four black hole statements and one galaxy statement, sA220, with sA220 having the lowest loading of the group. The black hole statements consist of sA237, sA238, sA235, and sA246, three of which appear in the group under the Black Holes factor label “Black Hole Properties,” in Table 9.10. That sA220, “all galaxies are spiral,” is mixed in with these statements suggests that the second group is associated with the misconception that all galaxies and black holes look the same. The fourth group consists of the same set of galaxy statements under the Galaxies factor label “Being in the Center,” in Table 9.4. The fifth group consists of the five highest-loading statements in the group under the Black Holes factor label “Science Fiction,” in Table 9.10.

11.11 Galaxies and cosmology

I performed PCA on the aggregate of statements about galaxies and cosmology, respectively with 12 and 5 statements in the associated topics of the ABI. Using the eigenvalue criterion, PCA supports four groups from the extraction of four factors, of which the first two are supported by PA. The first factor accounts for 27% of the total variance, and the first two factors cumulatively account for 37% of the total variance. Since there are only four factors in total, however, I will discuss all four resulting statement groups in the subsequent discussion. Table 11.11 presents the statements in all four groups.

Table 11.11 shows that, for the most part, misconceptions about galaxies are to be treated almost exclusively separate from misconceptions about cosmology. The first group consists of six galaxy statements and two cosmology statements. In the first group, the statements about galaxies are sA228, “all galaxies are the same in size and shape,” sA230, “the Milky Way is just stars — no gas and dust,” sA225, “there are only a few galaxies,” sA227, “we see all the stars that are in the Milky Way,” sA231, “new planets and stars don’t form today,” and sA220, “all galaxies are spiral.” Except

I			II			III			IV		
G	sA228	0.683	G	sA222	0.808	C	sA263	0.692	G	sA219	0.691
G	sA230	0.672	G	sA224	0.766	C	sA267	0.657	G	sA226	–0.479
G	sA225	0.596	G	sA221	0.680	C	sA256	0.616			
G	sA227	0.515	G	sA218	0.436						
G	sA231	0.476									
C	sA253	0.473									
C	sA262	0.464									
G	sA220	0.406									

Table 11.11. Reduced representation of statement factor loadings regarding galaxies and cosmology

for sA227, all of these galaxy statements appear in the first group of statements in my analysis of exclusively the Galaxies topic, as discussed in Section 9.1. The statements about cosmology are sA253, “gravity will eventually pull all the planets together,” and sA262, “the universe as a whole is static (unchanging).” Oddly enough, students who think that the universe is static also tend to think that gravity will pull all the planets together. Comprehensively, the first group of galaxy and cosmology statements pertains to the visual properties of galaxies, as if the galaxies have always been like how they appear today.

The second group consists of exclusively galaxy statements. The first three galaxy statements, sA222, sA224, and sA221, pertain to the nature of “Being in the Center,” as discussed in Section 9.1. PCA suggests that one should teach to misconceptions about “Being in the Center” separate from misconceptions about cosmology.

The third group consists of the remaining three cosmology statements: sA263, “astronomical ideas of mass, distance, and temperature of planets are all speculative,” sA267, “there is a center to the universe,” and sA256, “space is infinite.” These statements are also in the first group of statements in my analysis of the General Analysis topic of statements, as discussed in Section 10.7. PCA thus reinforces that misconceptions about the fundamental properties of the universe as a whole should be treated separately from misconceptions about galaxies.

The fourth group consists of two galaxy statements: sA219, “the solar system is not in the Milky Way (or any other) galaxy,” and sA226, “the galaxies are randomly distributed.” In my PCA of exclusively the galaxy topic, in Section 9.1, I mentioned that sA219 has a marginally significant negative loading on the same factor onto which sA226 loads the highest. Table 11.11 shows that when I include statements about cosmology and re-run the analysis, the anti-correlated nature between sA219 and sA226 is preserved.

11.12 Discussion

In this Chapter, I performed PCA on multiple aggregations of individual topics from the ABI. I showed that misconceptions associated with select statements are best unlearned by teaching in the context of the underlying theme in common with the objects of interest (e.g., atmospheres, scale sizes), rather than by teaching about each object independently. Some suggestions on how to teach students in the context of select misconceptions which they may harbor are as follows:

1. The size, temperature, and brightness of the Sun should be compared to other stars in the universe (Section 11.4).
2. The sizes of the Earth and Mars should be compared to other planets in the solar system (Section 11.6).
3. The lit and unlit sides of the Earth and the Moon should be used to illustrate their relative orientations and phases together (Section 11.7).
4. The changing length of the day on the Earth should be taught in conjunction with the direction that the Sun sets (Section 11.8).

In addition, when I include statements from multiple topics, PCA tends to preserve groups of misconceptions that would result from performing PCA on each topic separately. Some examples of preserved groups of misconceptions are as follows:

1. The Sun is the brightest, largest, and hottest of all the stars (Sections 10.6, 11.2, 11.8, and 11.9).
2. Black holes should appear like actual black holes and have no end to their lives (Sections 9.2 and 11.4).
3. Black holes permit space and time travel through them (Sections 9.2 and 11.10).
4. The Earth is at the center of the universe, and one can stand on it to watch objects in space move around it (Sections 10.5, 11.6, 11.8, and 11.9).
5. The universe has a center, goes on forever, and therefore is void of fundamental quantities such as mass, distance, and temperature (Sections 10.7 and 11.11).

CHAPTER 12

ITERATIVE PARALLEL ANALYSIS ON THE WHOLE DATA SET

In Chapter 11, I used PCA and PA to suggest groups of highly-correlated misconceptions, between and within ABI topics. First, I performed PCA on the aggregate of all of the statements from multiple topics. Then I determined the number of factors supported by PA and retained, for a subsequent analysis, only those statements which were grouped together within those factors. Some of these groups, each of which suggest the most logical subtopics to which one should teach, consist of statements from two or more sections of the ABI. My analysis thus suggests that there may be an underlying conceptual theme that ties together particular misconceptions from multiple topics in astronomy.

12.1 Iterative parallel analysis

To examine the likelihood of these underlying themes, I subject the student responses to all 215 statements of the ABI in a kind of iterative parallel analysis technique. The motivation is to retain, of the 215 statements, only a subset of the statements which consists of groups of the most highly inter-correlated statements throughout the ABI. The technique which I describe momentarily naturally reveals the most highly inter-correlated statements in the ABI and forms groups consisting of sets of the inter-correlated statements.

I use *iterative parallel analysis* to determine the most highly inter-correlated statements from the original set of 215 statements. First, I perform PCA on the entire data set as usual. Then I perform PA in the manner described in Section 6.3.3, and I retain only those statements which are grouped together within those factors supported by PA. Then I subject the reduced data set to iterative parallel analysis, which proceeds as follows:

1. Perform PCA and PA on the reduced data set.
2. Retain only those statements which are grouped together within those factors supported by PA.
3. Check the set of retained statements against the parent data set from Step 1.

One possibility is that the statement sets differ, in that the iteration produces a further-reduced statement set. If the statement set reduces further, then I perform iterative parallel analysis on the reduced statement set. If, however, the statement sets are the same, then the set cannot be reduced further. Of this final reduced statement set, I may examine the statement groups supported by PA. Figure 12.1 presents a flow chart of the iterative parallel analysis technique.

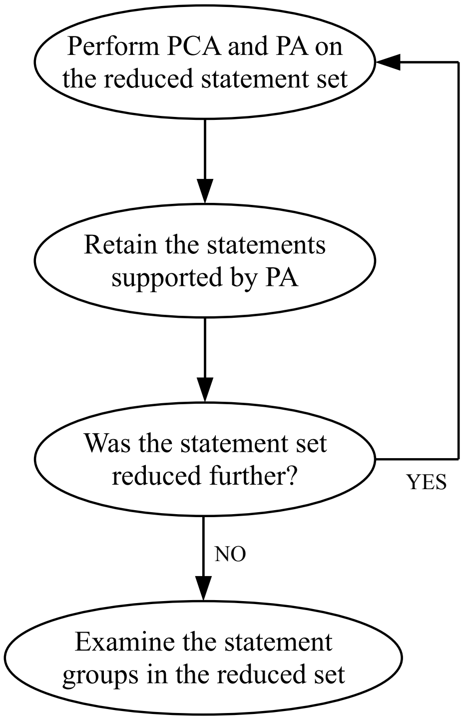


Figure 12.1. Iterative parallel analysis flow chart

12.2 Results

I performed iterative parallel analysis, as discussed in Section 12.1, on the original set of 215 statements from the ABI. After five iterations, I retained 27 statements divided into four groups. Figure 12.2 shows the scree plot that results from performing iterative parallel analysis. To confirm that the cut-off at the fourth component is not ambiguous, I ran 10 additional parallel analysis sequences of 500 repetitions, from which I found that the observed eigenvalue, of 1.33, is more than 10σ away from the distribution of eigenvalues derived from the additional parallel analysis sequences at the fourth component. Figure 12.2 further shows that extraction of four factors is supported by the criteria by Kaiser (1960) and Cattell (1966), and passes the parallel analysis test Horn (1965), which makes the choice of number of factors to extract unambiguous. Table 12.1 presents the statements in the four groups. The reader may refer to Table 11.1 (page 193) for a list of symbols that represent the topics of the ABI. The reader may also refer to Appendix A for a list of statements in the ABI.

	I			II			III			IV	
⊕	sA115	0.760	⊙	sA189	0.795	♂	sA168	0.755	G	sA222	0.754
⊕	sA116	0.725	⊙	sA198	0.753	♂	sA167	0.731	G	sA224	0.653
⊕	sA133	0.668	⊙	sA209	0.751	♂	sA161	0.703	G	sA221	0.636
§	sA51	0.645	⊙	sA190	0.717	⊕	sA157	0.643	⊕	sA150	0.621
⊕	sA160	0.569	⊙	sA180	0.632	§	sA53	0.491	⊙	sA181	0.521
◑	sA77	0.534	★	sA17	0.539	♂	sA169	0.476	G	sA218	0.463
◑	sA92	0.447									
◑	sA87	0.433									
♂	sA166	0.416									

Table 12.1. Reduced representation of all statement factor loadings after performing iterative parallel analysis

Table 12.1 shows that the statements that have the very highest-loading inter-item correlations are contained to within a single topic of the ABI. The first group consists of four Earth statement sA115, “Earth is at the center of the universe,” sA116, “Earth is the biggest planet,” sA133, “the sun orbits the Earth,” and sA160, “the Earth

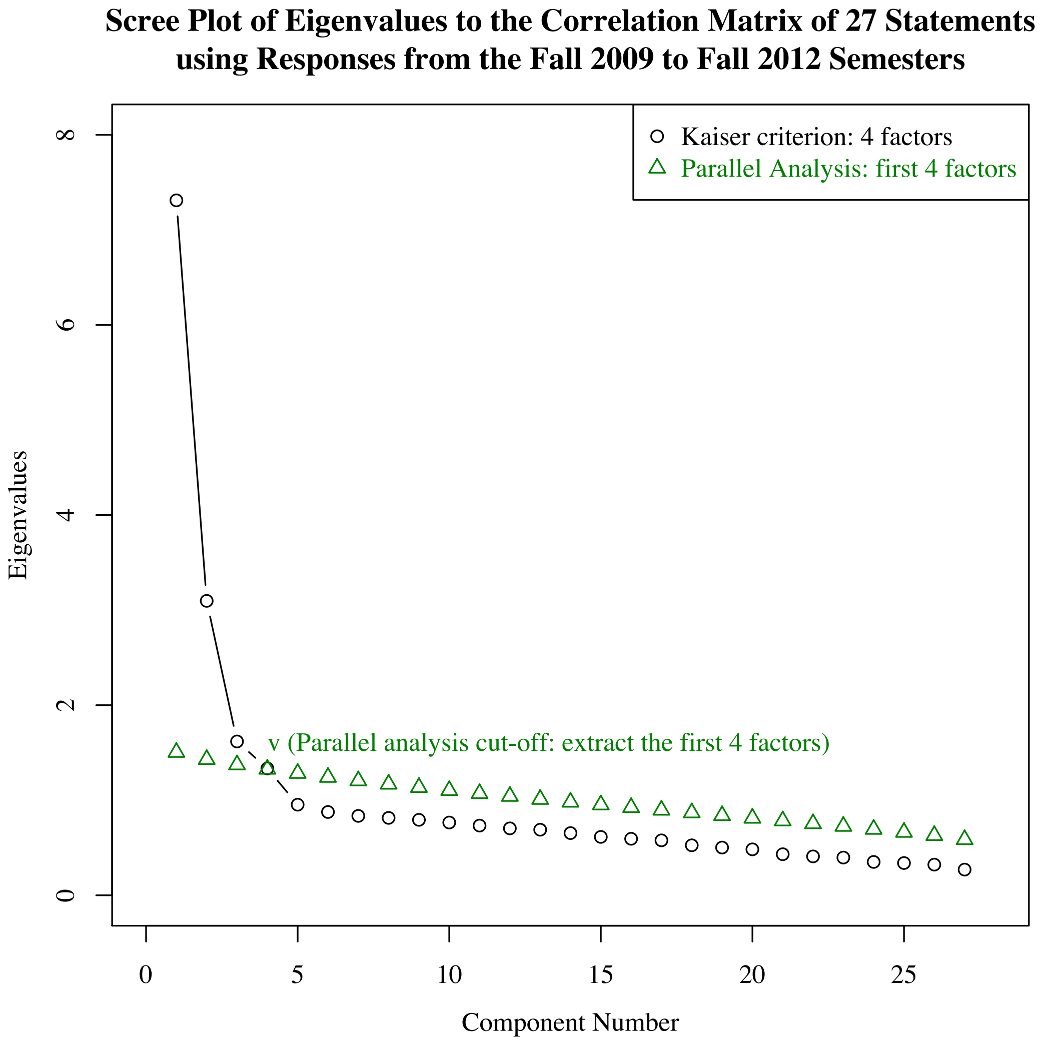


Figure 12.2. Scree plot from performing iterative parallel analysis

is flat." These four statements were previously found to be highly inter-correlated and pertain, in part, to the overarching misconception that the Earth is at the center, and that one can stand on the Earth to watch objects in space move around it (Sections 10.5, 11.6, 11.8, and 11.9). The statement about the solar system is sA51, "planets revolve around the Earth," which further reinforces this misconception. The statements about the Moon are sA77, "there is only one moon — ours," sA92, "the Moon has a smooth surface," and sA87, "the Moon has seas and oceans of water." Both sA160 and sA92 have high inter-correlations and pertain to the nature of surfaces being flat. Fortunately, iterative parallel analysis actually *preserves* the reoccurring theme, that one stands at the center, on the Earth, to watch objects go around it.

In Table 12.1, the second group consists of five Sun statements, sA189, sA190, sA198, sA209, and sA180, which were previously identified as statements in which the Sun is represented as the brightest, largest, and hottest of all the stars (Sections 10.6, 11.2, 11.8, and 11.9). Statement sA17, "all stars are smaller than the Sun," inter-correlates with these five Sun statements. Once again, iterative parallel analysis actually *preserves* the reoccurring theme that the Sun is the brightest, largest, and hottest of all the stars.

The third group consists of four Mars statements, one Earth statement, one Solar System statement, and one Earth statement. In the third group, the statements about Mars are sA168, "there are Lowellian canals on Mars built by intelligent beings," sA167, "life, when it did exist on Mars, was quite advanced," sA161, "Mars is green (from plant life)," and sA169, "Mars is Hot because it is red ... Mars — god of fire." The statement about the Earth is sA157, "Earth has a second moon that only comes around once in awhile — 'once in a blue moon'," a statement which was previous found to correlate with statements about life on Mars, as discussed in Section 11.6. The statement about the Solar System is sA53, "Pluto is a large, jovian (Jupiter-like) planet." Comprehensively, the statements in the second group seem to pertain to how students would describe the surfaces of the Moon, the Sun, Mars, and Pluto. Iterative parallel analysis thus suggests that students compartmentalize their recognitions of the Moon, the Sun, and Mars together by associating them with descriptions of how they would appear from the Earth.

The fourth group consists of four Galaxy statements, one Earth statement, and one Sun statement. In the fourth group, the first three galaxy statements relate to “Being in the Center” of the Milky Way or of the universe, and are the same as those in the second group of Table 9.3. It should thus not be too surprising that sA150, “the Earth is in the middle of the Milky Way galaxy,” strongly inter-correlates with these galaxy statements. Incidentally, students who endorse these statements also think that the Sun does not move through space, and that there is only one galaxy — ours.

12.3 Discussion

In Section 12.2, I presented the results of performing iterative parallel analysis on the entire data set. I showed that the inclusion of all 215 statements in the inventory naturally reduces to four groups of highly inter-correlated statements. Each of the four groups consists of statements whose underlying conceptual themes are supported by previous analyses on smaller portions of the ABI, either one topic at a time (e.g., only Sun statements), or with two or three topics at a time (e.g., Moon, Earth, and Sun statements). My iterative parallel analysis technique thus tends to reaffirm the existence of these underlying themes, which supports the use of PCA and PA to identify groups of correlated statements with an underlying misconception-based theme. In order of decreasing factor variance, the underlying themes from the four groups are:

1. the Earth is at the center, and one can stand on the Earth to watch objects in space move around it,
2. the Sun is represented as the brightest, largest, and hottest of all the stars,

3. recognition of the Moon, the Sun, and Mars is associated with descriptions of how they would appear from the Earth, and
4. we are fixed at the center of the Milky Way and/or the universe.

CHAPTER 13

ITEM RESPONSE THEORY

In Chapters 8-10, I performed PCA on each topic of the ABI and, within each topic, determined groups of correlated misconceptions to which one can teach. In Chapter 11, I performed PCA and PA on aggregates of two or more topics of the ABI to examine groups of underlying themes (e.g., scale sizes of the planets) that link misconceptions throughout various topics of the ABI. In Chapter 12, I reduced the entire set of 215 statements from the ABI using iterations of PCA and PA to determine four correlated sets of misconceptions throughout various topics of the ABI. In this Chapter, I introduce the elements of item response theory, the methodology of which provides relative measures of the difficulties of unlearning the misconceptions associated with statements in each of the groups.

13.1 Limitations of factor analysis

My application of factor analysis to the ABI data is limited in four fundamental ways. First, *the application of factor analysis is limited to the set of statements that one chooses to call a topic*. The topics of the ABI were chosen and organized by NFC from his experiences as an instructor in AST 109. Given my analysis of statement presentation order, in Section 3.3, however, the proximity of neighboring statements had no significant influence on student responses, so the statements could be resorted into different topics. Second, factor analysis is *not an automatic script* that tells the researcher how to sequence topics in the introductory astronomy course. As taught here at the University of Maine, topics in the introductory astronomy course are sequenced in such a way that knowledge of some topics requires the knowledge of one or more previous topics. Third, *factor analysis does not provide information about the difficulty of the items*. Factor analysis uses other correlations or covariances between any two items, to determine the extent to which students will consistently endorse or reject misconceptions associated with statements in a group. The factor loadings, however, do not describe the actual proportion of students who endorse the misconceptions. In other words, factor analysis does not provide statistics on the scores themselves. And fourth, while factor analysis groups statements together, the analysis provides no recommendation on specifically *how to teach in the context of the associated misconceptions individually*.

While factor analysis thus provides no recommendation for how to sequence topics in astronomy, the optimal sequence to teach items *within groups* can be determined. Since factor analysis can determine groups of the most inter-correlated items within a topic, one can use these groups individually to analyze the proportions of the actual scores on the constituent items. In order to determine the most logical sequence of the items, I analyze the ABI scores to the statements in each group, one group at a time. The analysis follows by subjecting the ABI data to a fit of the data to a model that follows from the methodology of item response theory, which I discuss in Section 13.3.

13.2 Classical test theory vs. item response theory

As noted by Morizot, Ainsworth, and Reise (2007) and Wallace and Bailey (2010), the common practice to analyze scores (e.g., the three response codes in Table 2.4) in psychometric analysis is to use classical test theory. In classical test theory, one can compare the difference in the observed vs. true subject score, or the observed score variation vs. the true score variation. The reliability of classical test theory, however, depends on parameters that are strongly influenced by the sample. For example, in classical test theory, the difficulty of an item is calculated based on the proportion of students who endorse the incorrect response. In classical test theory, one would conclude that the hypothetical score of the students, after removing effects due to measurement errors (e.g., a person's mood) and systematic errors (e.g., room lighting, keeping track of question number), depends on their responses to the items.

An alternative is to apply the methodology of item response theory (IRT), which establishes relative item difficulties independent of sample characteristics, a feature in IRT known as *parameter invariance*. For the purposes of analyzing responses to the ABI, the item *difficulty* is the degree to which a misconception is expected to persist with the students in the sample. As Morizot et al. point out, IRT is a statistical theory that measures the extent to which some particular latent trait of a subject influences their scores. A *latent trait* is an inferred (or unobserved) variable that represents some attribute of a subject, such as the overall tendency to endorse misconceptions, and the value of that attribute is directly expected to affect the subject's responses. The value of the latent trait is designated *ability* in the literature. IRT is also capable of clarifying the extent of discrimination between two subject groups, that is, "to differentiate between individuals at different trait levels" (p. 409), or at different abilities. As Kline (2005) and Funk and Rogge (2007) note, given a large sample size for a group of well-correlated items, the standard error of the mean per item converges more rapidly in IRT than in classical test theory. In other words, in the limit of large sample sizes, IRT analyses can give results that are more precise (have lower standard errors) than a corresponding analysis of the data using a classical test theory model. I have thus elected to use IRT instead of classical theory to analyze student responses to the ABI.

13.3 The methodology of item response theory

A core assumption of IRT methodology is the assumption of *unidimensionality* among a set of items. As stated by Morizot et al. (2007), unidimensionality is the property that "a single latent trait accounts for all the common variance among item responses" (p. 413) In IRT, the items in the set are *locally independent* if, as Morizot et al. continues, “once the latent trait is controlled for, there are no significant correlations left among the items.” For example, in an analysis of responses to the ABI, in which the latent trait is the tendency to endorse misconceptions, only misconception endorsement significantly correlates with the raw ABI statement responses. Morizot et al. (2007) further note that in practice, the unidimensionality requirement is rarely satisfied. Instead, one is often interested in whether or not the set of items is “unidimensional enough,” which is to say that there are no other latent traits that significantly impact the data.

In Section 7.3, I declared that I would use data from the aggregate of exclusively the Fall 2009 through Fall 2012 semesters. As I explained in Section 7.3, if I include data from the Spring 2013 and Fall 2013 semesters, the groups that result from performing PCA would not represent groups of *correlated misconceptions*. Instead, the groups would comprise “sometimes misconceptions” statements, rather than always being phrased as misconceptions, as in the first four semesters. The act of mixing sometimes-true and always-false statements inconsistently among the groups raises concerns about the dimensionality of the items, in that there are now two latent traits: “misconception endorsement,” and “fact endorsement.” The use of a multidimensional IRT model to examine the quantitative statistical differences between these two latent traits is outside of the scope of my dissertation.

The groups of inter-correlated statements produced by performing PCA, as presented in Chapters 8-10, represent example item sets, as the items in each set are reasonably well inter-correlated. The use of PCA with orthogonal rotation (refer to the discussion of factor rotations in Section 6.1.2) minimizes the severity of any violation of local independence. As Edelen and Reeve (2007) examine, while violations of local independence may still occur, the researchers note that “IRT applications are often robust to violations of local dependence,” and “in general, it is preferable to retain the full set of items, especially if they comprise a commonly used existing scale” (p. 11).

My ABI data is coded consistently for all 215 statements in the ABI. Since I have already subjected these data to a factor analysis to generate groups of statements, I conclude that a unidimensional IRT model is sufficient to generate item parameters within each group of statements.

With regard to sample size, Morizot et al. note that the items themselves should have enough common variance to give reasonably unbiased estimates of item difficulty. They state that in general, an unbiased analysis for dichotomously-scored items (those with two possible response codes, e.g., 0 or 1) may have as few as 100 subjects, whereas 5-point response formats require a sample size of at least 500 subjects. For my analysis of the ABI, I have three possible response codes, with a sample size of $N = 401$ for the Fall 2009 to Fall 2012 semesters under consideration, so I can reasonably assume that my sample size is large enough for me to run IRT tests on the statements in each group, one group at a time.

13.4 Item response theory models

13.4.1 The 1 and 2 parameter logistic models

That student responses are tabulated based on the frequency of the scores in Table 2.4 allows me to construct directly a logistic curve model for the probability distribution of each ABI score. Logistic curves indicate the probability of obtaining at least a certain score in respondents, as a function of the latent trait of the subjects as a whole (McDonald, 1999; Morizot et al., 2007; Wallace & Bailey, 2010, and references therein). Figure 13.1 presents an example plot of a logistic curve. The features of the plot are explained in the subsequent paragraph.

Since part of my analysis involves polytomous scoring (with more than two possible responses per item), I first describe the construction of the logistic curves for dichotomous scoring, then I generalize to polytomous scoring using the graded response model. Hereafter, I follow the construction of the 1PL and 2PL models as

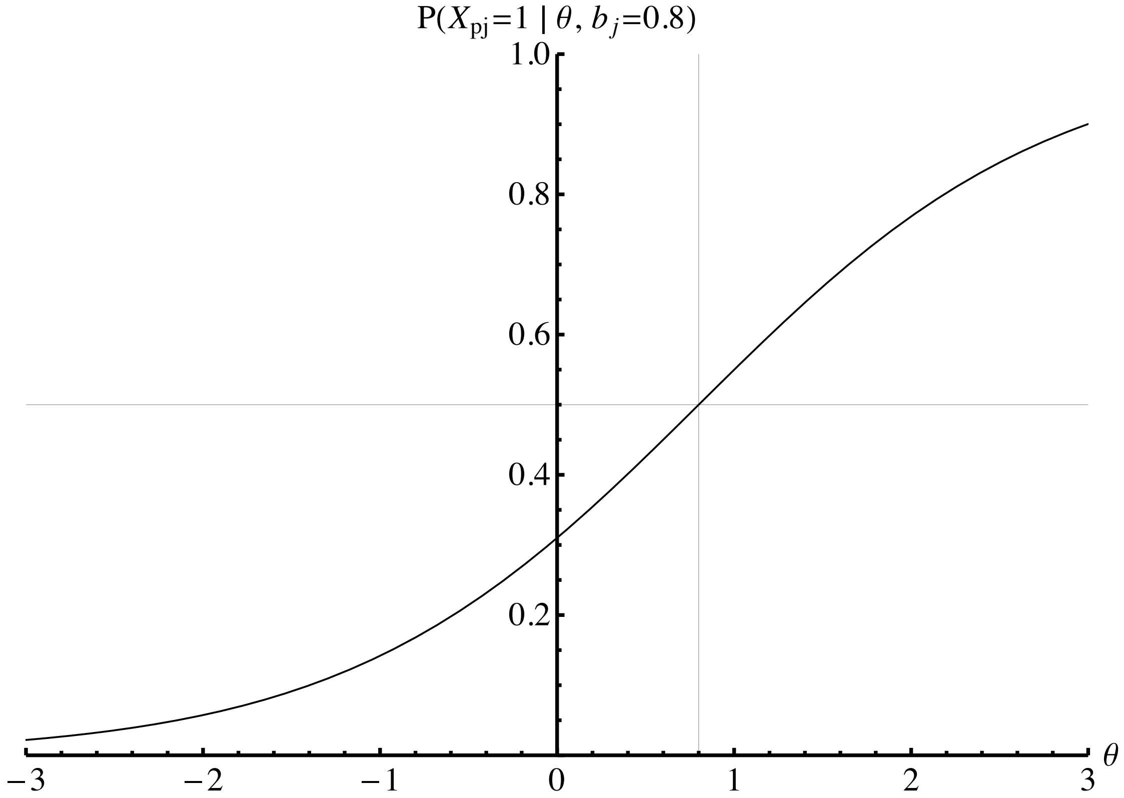


Figure 13.1. Graph of an item characteristic curve with the parameter $b_j = 0.8$

presented by (Morizot et al., 2007). In the one-parameter logistic (1PL) model, or Rasch model, the probability for subject p whose ability level is θ to obtain a score X_{pj} (typically taken to be 1 for a correct answer and 0 for an incorrect answer) for a dichotomously scored item j of difficulty (or "item parameter") b_j is given by

$$P(X_{pj} = 1|\theta, b_j) = (1 + e^{-(\theta - b_j)})^{-1}, \tag{13.1}$$

where b_j is the value of θ at which the probability of obtaining the score is 50%, or 0.5. The plot of $P(X_{pj} = 1|\theta, b_j)$ vs. θ is called the item characteristic curve (ICC). The ICC in Figure 13.1 has the parameter $b_j = 0.8$, such that at the ability level $\theta = 0.8$, subject p has a 50% chance of responding with the correct answer. The θ axis is scaled on a z-score metric such that its mean and standard deviation are respectively 0 and 1.

The 1PL model is used to determine the probability of responding with the correct answer, as a function of some latent trait of the subjects. An example of a latent trait used by Thorpe et al. (2007) in their study of common self-evaluative irrational beliefs is the subjects' degree of irrational self-evaluation, in which case "irrationality" would be the label for the θ axis. For my study, I am interested in how the probability of responding with a particular degree of misconception retainment, given by the response codes (1, 2, 3) in Table 2.4, changes with the statements in each of my statement groups. The codes approximately model relative times in one's life as to when one disabuses oneself of a misconception, or if one retains the misconception well into adulthood. Hence, I will refer to the ability (θ) axis as a scale of *misconception endorsement*, which represents the characteristic latent trait of the students. Recall that the θ axis scales values of misconception endorsement, taken from the ABI response data, on a z-score metric, with mean of 0 and standard deviation of 1. So, negative values of misconception endorsement refer to an increasing tendency for students to disabuse themselves of misconceptions, while positive values of misconception endorsement refer to an increasing tendency for students to retain misconceptions.

In the 1PL model, the shape of the probability distribution is the same for each item in a group. For my analysis of the topics of the ABI, I note, for example, that some items are expected to discriminate more sharply between higher and lower-achieving students than other items. For example, as I show in Section 14.3.1, sA234, "black holes really don't exist," is less able to discriminate between higher and lower-achieving students than other statements in the group, which suggests that an additional parameter should be incorporated to describe the discrimination tendency of individual statements.

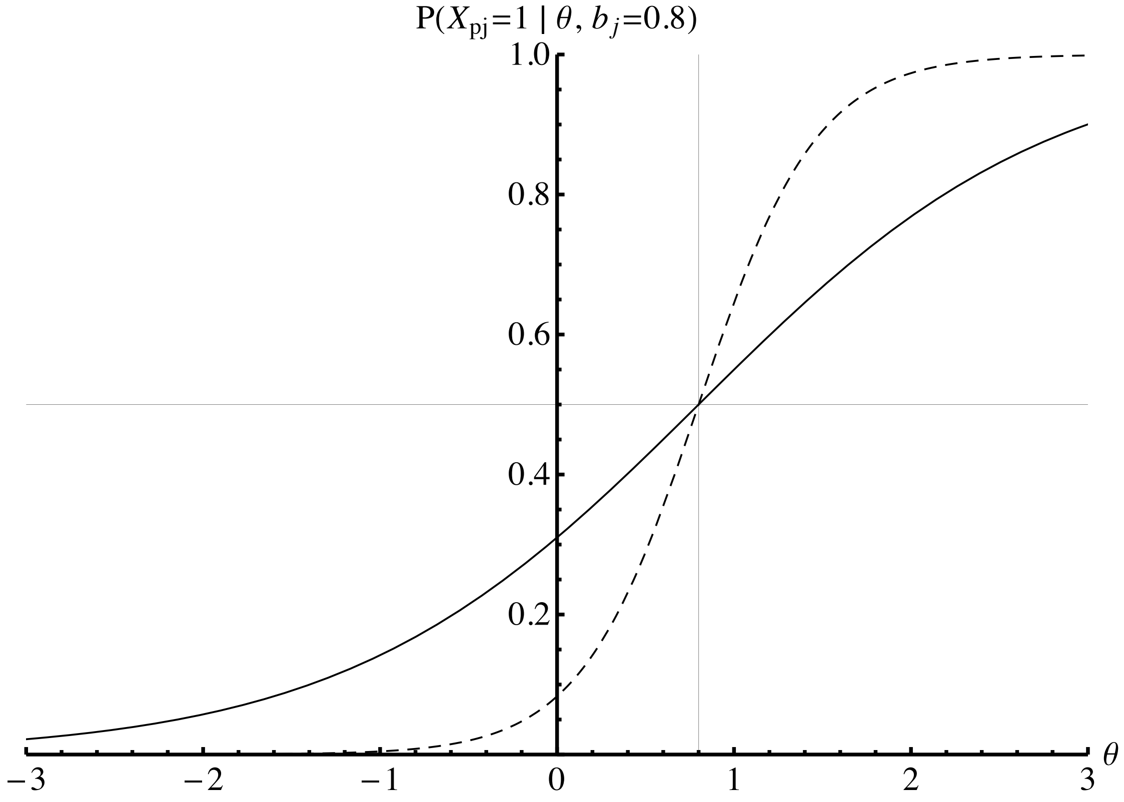


Figure 13.2. Graph of two item characteristic curves on a plot with mean of 0 and standard deviation of 1, with the parameter $b_j = 0.8$, but with different discriminations (solid line, $a = 1$; dotted line, $a = 3$)

In order to incorporate discrimination among the grouped items, I use a two-parameter logistic (2PL) model, as suggested by Wallace and Bailey (2010), with a discrimination parameter a_j that measures the steepness of the ICC. The 2PL model is written as

$$P(X_{pj} = 1|\theta, a_j, b_j) = (1 + e^{-a_j(\theta - b_j)})^{-1}. \tag{13.2}$$

A high value for a_j corresponds to high item discrimination, which represents a strong division between higher and lower-achieving subjects for the item. In Figure 13.2, the steepness, or discrimination, at $b_j = 0.8$ of the second ICC (dashed curve) is three times higher than that of the previous item (solid curve). For the 2PL model, the discrimination parameter can be thought of as being inversely related to the standard error of b_j, since item characteristic curves are also the steepest at the value b_j and so are the most able to discriminate between students who respond correctly and students who respond incorrectly. Values of $a_j \to 0$ correspond to a broader mixing of subject misconception endorsement levels. Items whose discrimination parameters $a_j < 1$ are said to have a relatively low ability to discriminate between higher and lower-achieving subjects (Baker, 2001).

Models that introduce additional parameters exist in the literature. For example, there exists a three-parameter logistic (3PL) model, which incorporates a guessing parameter c_j for right or wrong answers. The 3PL model is not applicable to my study, because the ABI is not a multiple-choice test, so $c_j = 0$. Edelen and Reeve (2007) briefly mention that a four-parameter logistic model provides estimates of upper asymptotes to the item characteristic curve. For my analysis of the ABI data, however, I am most interested in the discrimination and item parameters a_j and b_j.

13.4.2 The graded response model

While the 1PL and 2PL models are used in IRT for dichotomous scoring, a polytomous model allows one to analyze more than two responses per item at a time. Hence, there will be more than one "characteristic curve" to plot the probability of responding with a particular score. The colloquialism in the literature, as noted by Reeve and Fayers (2005), is to refer to such a curve as a characteristic response curve (CRC). I will hereafter refer to the curves as CRCs.

Statistical analysis in the ABI involves polytomous scoring. To perform IRT analyses on polytomous scoring, one may choose from a number of models (Edelen & Reeve, 2007). Of these models, the graded response model (GRM), first introduced by Samejima (1969), is most relevant to my study and naturally extends the 2PL model to the case of polytomous scoring. According to Embretson and Reise (2000), in the GRM, "the items need not have the same number of response categories; no complications arise in item parameter estimation or the subsequent parameter interpretation as a result of a measure having items with different response formats" (pp. 97-98). The GRM preserves the order of the response codes and is adequate as long as the response codes follow a consistent rating structure, which they do in the ABI. All of the items in the ABI are misconception-based statements, and the ABI data for each item follows a consistent rating structure. Had the ABI data not been coded consistently, an alternative model, such as the partial credit model (Masters, 1982), would be necessary. In the partial credit model, one first establishes a special code structure to score each item, then models the probability of responding with each code to each item. Therefore, I have elected to use the GRM for my analysis. In the GRM, the total probability of any response is normalized to 1, that is,

$$\sum_{k=1}^{K} P_{jk}(\theta) = 1 \tag{13.3}$$

for K scores. One can thus plot all of the CRCs on the same graph, as in Figure 13.3, where now b_j is a vector that represents the item parameter for each CRC. Additional clarification on the GRM is presented in the subsequent paragraph.

Because of probability conservation, $K - 1$ CRCs have a defined parameter, and I follow the convention introduced by Samejima to chose the locations of all but the first CRCs. Hence, in Figure 13.3, the second through fifth CRCs (in order: dashed, dot-dashed, thick-dotted, and thick-long-dashed), have the respective parameters $b_j = (-2.0, 0.0, 1.5, 2.5)$. Because of probability conservation, the first CRC (left-most, dotted) has no item parameter. I specifically note that I now require that b have two subscripts, one for item j, and one for the particular score k, so that the item parameter for item j and score k is b_{jk}. In Figure 13.3, if the response codes are in the range $1 - 5$, then $b_{j2} = -2.0$, $b_{j3} = 0.0$ $b_{j4} = 1.5$, and $b_{j5} = 2.5$. The *vector* $b_j = (-2.0, 0.0, 1.5, 2.5)$ then represents the set of these item parameters.

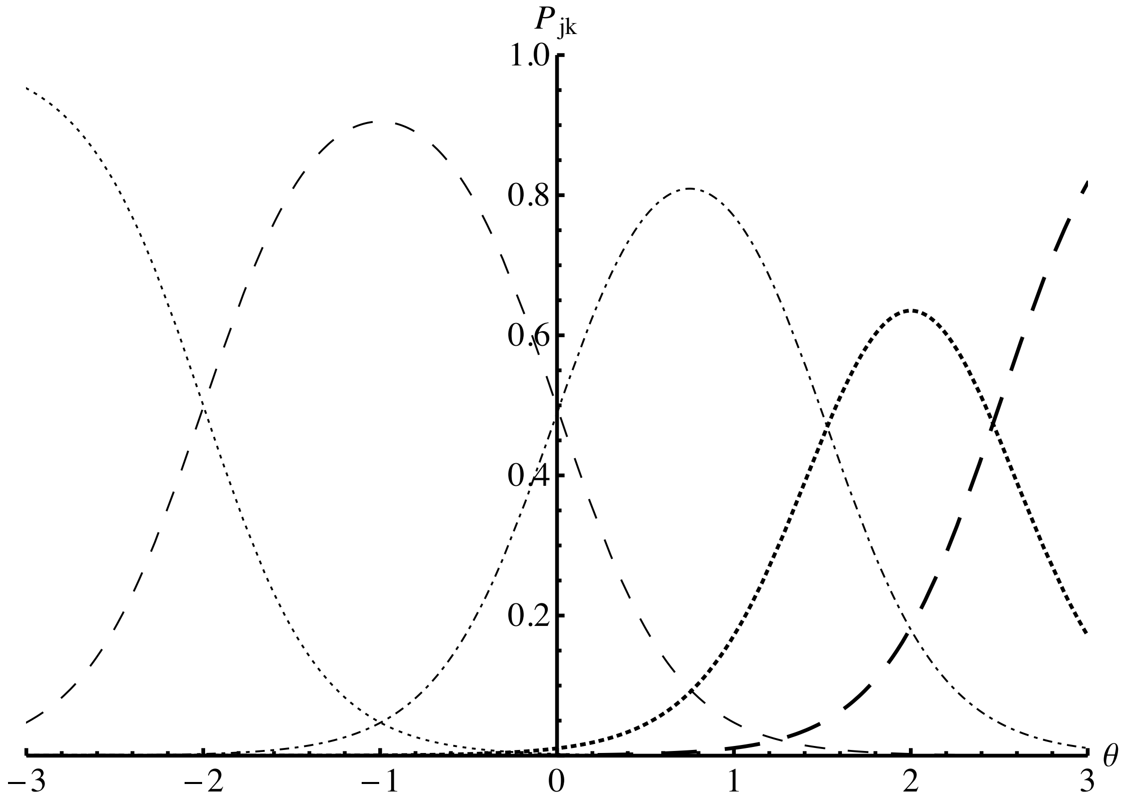


Figure 13.3. Graph of five characteristic response curves using the graded response model, with $b_j = (-2.0, 0.0, 1.5, 2.5)$

Following the construction by Samejima (1969), in which k goes from 1 to K, the 2PL function for the highest score, P_{jK}, is given by

$$P_{jK}(\theta) = (1 + e^{-a_j(\theta - b_{jK})})^{-1}. \tag{13.4}$$

The 2PL for the $K - 1^{\text{th}}$ score is then

$$P_{j,K-1}(\theta) = (1 + e^{-a_j(\theta - b_{j,K-1})})^{-1} - P_{jK}(\theta). \tag{13.5}$$

To conserve probability, the 2PL for the $K - 2$th score is then

$$P_{j,K-2}(\theta) = (1 + e^{-a_j(\theta - b_{j,K-1})})^{-1} - P_{j,K-1}(\theta) - P_{jK}(\theta). \tag{13.6}$$

This pattern continues until one reaches the lowest score, which is

$$P_{j1}(\theta) = 1 - P_{j2}(\theta) - \cdots - P_{jK}(\theta). \tag{13.7}$$

Samejima (1969) introduces the notation

$$P^{+}_{jk}(\theta) = P_{j,k+1} + P_{j,k+2} + \cdots + P_{jK}, \tag{13.8}$$

where $P^{+}_{jk}(\theta)$ is the sum of all CRCs from $P_{j,k+1}$ up to P_{jK}. Comprehensively the probability of responding with score k in the GRM for each item j is

$$P_{jk}(\theta) = \begin{cases} 1 - P^{+}_{jk}(\theta) & \text{for} \quad k = 1, \\ (1 + e^{-a_j(\theta - b_{jk})})^{-1} - P^{+}_{jk}(\theta) & \text{for} \quad 2 \leq k \leq K - 1, \\ (1 + e^{-a_j(\theta - b_{jK})})^{-1} & \text{for} \quad k = K. \end{cases} \tag{13.9}$$

One constraint of the GRM is that of a single discrimination to describe the relative steepness of each CRC. Consider, for example, the CRC plots in Figure 13.3. Software that uses the GRM performs a “best fit” to the data by utilizing a single discrimination parameter a_j to describe the relative sharpness in transitions between pairs of two neighboring scores, which, as I explain in Section 13.6, tend to occur near the locations of the item parameters b_{jk}. The GRM altogether ignores that, along the θ axis, the transition from score $k - 1$ to score k, for example, may in practice be rather gradual, whereas the transition from score k to score $k + 1$ may be relatively sudden. The effect is that, near $\theta = b_{jk}$, the probability of responding with a particular score may be misestimated. The severity of the violation depends on the difference between the actual nature of the transition and its modeled steepness, as reflected by the value of a_j. The actual nature of the transitions, however, is often unknown in practice, and may be extremely challenging to quantify beyond the usual implementation of the discrimination parameter a_j. In other words, while an interested researcher could construct a model that uses two or more discrimination parameters, per item, to improve the fit of the CRCs, the increased complexity of the model also comes with no known way of testing the validity of the additional discrimination parameters, which makes such a more complex model rather impractical. Hence, I have elected to restrict my analysis to that of a single discrimination parameter.

13.5 Item information

In Fisher information theory, the item information is the statistical variance of the score (Lehmann & Casella, 1998). Conceptually, information is a relative measure of the reliablility of the value of a CRC at θ, or how well each score is being estimated at θ. Hence, in IRT literature, one typically refers to an item as being "most informative" (or "best estimating" each score) at the location on the θ axis where the item information curve peaks. Away from these peaks, the scores are not estimated very well, and so the reliability of the CRC values for the score decreases with information (Baker, 2001). Extending Fisher information theory to polytomously-scored items, the item information curve for item j is given (Chajewski & Lewis, 2009) by

$$I(\theta) = \sum_{k=1}^{K} \frac{1}{P_{jk}(\theta)} \left(\frac{dP_{jk}(\theta)}{d\theta} \right)^2 . \tag{13.10}$$

Note that $I(\theta)$ depends on the square of the derivative of the CRCs.

Referring to Section 13.4.1, the role of the discrimination parameter a_j is to discriminate between students who reject a misconception and students who endorse it. A higher value of a_j leads to steeper CRCs, as given by Equation (13.2). Steeper CRCs increase the value of $dP_{jk}(\theta)/d\theta$ at the location on the θ axis where the CRC is the steepest. Hence, given two items with equivalent sets of item parameters b_j, the item with the higher value of a_j is more informative than the item with the lower value of a_j. Not surprisingly, a lower value of a_j tends to produce less steep CRCs and lower total information. As in the plot of CRCs in Figure 14.3 (page 257), the discrimination

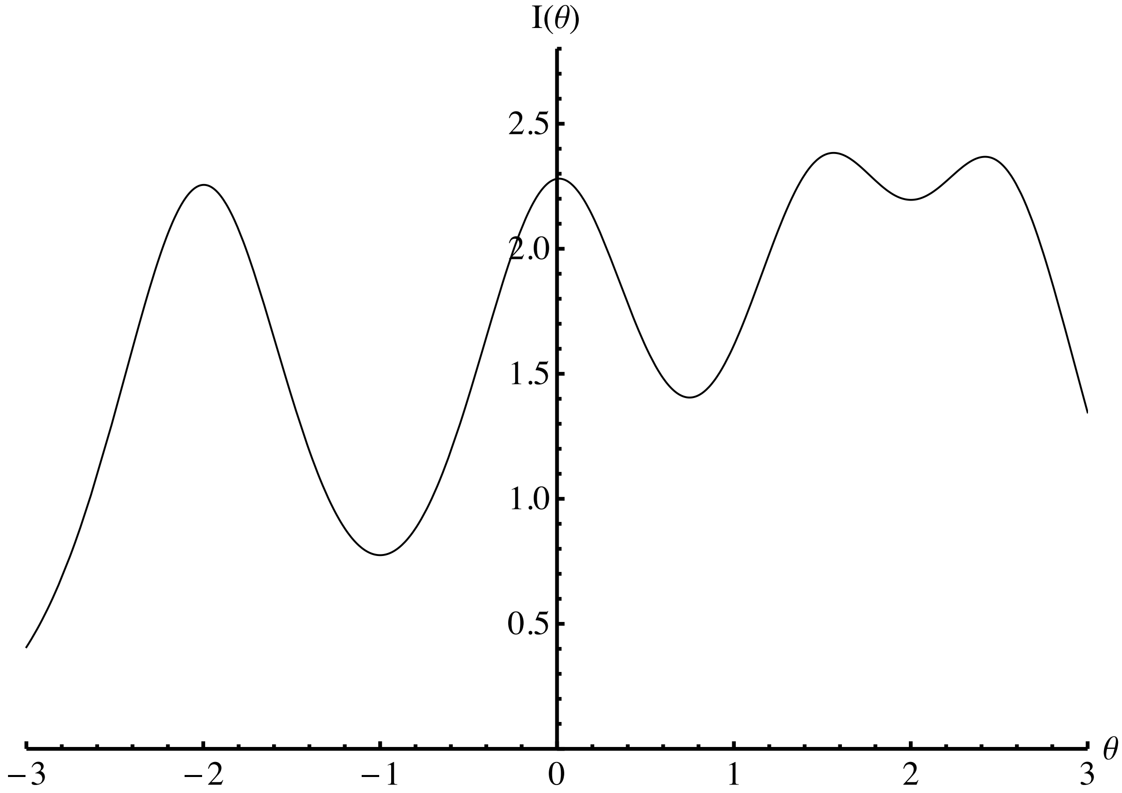


Figure 13.4. Graph of the total information for the five characteristics curves in Figure 13.3

parameter for each of the CRCs is $a_j = 0.87$, which corresponds to an item with a low ability to discriminate between students who endorse misconceptions vs. students who reject misconceptions. The consequence is that the total information $I(\theta)$ also tends to be lower overall. Figure 13.4 shows the total information $I(\theta)$ for the example in Figure 13.3.

Local information maxima are obtained by finding the abilities θ_s that satisfy

$$\left.\frac{dI(\theta)}{d\theta}\right|_{\theta=\theta_s} = \sum_{k=1}^{K} \frac{d}{d\theta}\left(\frac{1}{P_{jk}(\theta)}\left(\frac{dP_{jk}(\theta)}{d\theta}\right)\right)\Bigg|_{\theta=\theta_s} = 0, \tag{13.11}$$

where θ_s marks the location on the θ axis of a transition from one score to another (that is, at the location where neighboring CRCs intersect), and where

$$\left.\frac{d^2 I(\theta)}{d\theta^2}\right|_{\theta=\theta_s} < 0. \tag{13.12}$$

A quick approximation for θ_s, at the location where two neighboring CRCs with scores k and $k+1$ intersect can be found by estimating the coordinates of the information peaks. My estimation assumes that the contribution to $I(\theta)$ from all other CRCs is negligible; in other words, for all other scores n, $P_{jn}(\theta) \approx 0$. If, in Eqn. (13.11), I assume that two neighboring CRCs overlap with $P_{jn} \approx 0$, then

$$\frac{d}{d\theta}\left(\frac{1}{P_{jk}(\theta)}\left(\frac{dP_{jk}(\theta)}{d\theta}\right)^2\right) = -\frac{d}{d\theta}\left(\frac{1}{P_{j,k+1}(\theta)}\left(\frac{dP_{j,k+1}(\theta)}{d\theta}\right)^2\right). \tag{13.13}$$

After taking the derivative and setting $\theta = \theta_s$, one has, on the left-hand side,

$$\frac{dP_{jk}(\theta_s)}{d\theta_s}\left(\frac{2}{P_{jk}(\theta_s)}\frac{d^2 P_{jk}(\theta_s)}{d\theta_s^2} - \frac{1}{P_{jk}^2(\theta_s)}\left(\frac{dP_{jk}(\theta_s)}{d\theta_s}\right)^2\right)$$

and, on the right-hand side,

$$-\frac{dP_{j,k+1}(\theta_s)}{d\theta_s}\left(\frac{2}{P_{j,k+1}(\theta_s)}\frac{d^2 P_{j,k+1}(\theta_s)}{d\theta_s^2} - \frac{1}{P_{j,k+1}^2(\theta_s)}\left(\frac{dP_{j,k+1}(\theta_s)}{d\theta_s}\right)^2\right).$$

The only way for the left and right sides of the equation to be equal is if

$$P_{jk}(\theta_s) = P_{j,k+1}(\theta_s), \quad \text{and} \quad \frac{dP_{jk}(\theta_s)}{d\theta_s} = -\frac{dP_{j,k+1}(\theta_s)}{d\theta_s}. \tag{13.14}$$

But the identification $P_{jk}(\theta_s) = P_{j,k+1}(\theta_s)$ is also where two CRCs intersect. Therefore, if the contribution to $I(\theta)$ from each of the other CRCs is negligible, then *maximum information is obtained at approximately the intersection of CRCs, which is also where the magnitude of the CRC slopes are equivalent to each other.*

For example, consider a model with three possible scores ($K = 3$), and the item parameters $a = 1.0$, and $b_{jk} = (-1.0, 1.5)$. Consider the transition from score $k = 2$ to score $k = 3$. According to the GRM, the peak of the information curve corresponding to this transition occurs at $\theta = 1.02$, and the transition occurs at $\theta_s = 1.32$. At the location of the transition, $P_{j2}(\theta_s) = P_{j3}(\theta_s) = 45.5\%$.

As another example, consider a model with a higher discrimination parameter $a = 1.6$, and $b_{jk} = (-1.0, 1.5)$. Again, consider the transition from score $k = 2$ to score $k = 3$. According to the GRM, the peak of the information curve corresponding to this transition occurs at $\theta = 1.46$, and the transition occurs at $\theta_s = 1.48$. At the location of the transition, $P_{j2}(\theta_s) = P_{j3}(\theta_s) = 49.0\%$.

Note that for polytomous scoring, up to $K - 1$ peaks at the solutions θ_s may be clearly identified from the shape of $I(\theta)$. When the parameters b_{jk} for two neighboring CRCs match more closely, their information peaks can either form a plateau or converge into a single peak in the plot of $I(\theta)$ vs. θ. To illustrate the appearance of these plots, Figure 13.5 presents a plot of the five characteristic response curves in Figure 13.3, except with $b_j = (-2.0, 0.0, 1.5, 2.3)$.

For example, if, instead of my example parameters being $b_j = (-2.0, 0.0, 1.5, 2.5)$, I had $b_j = (-2.0, 0.0, 1.5, 2.3)$, the total information curve would resemble that in Figure 13.5, in which the right-most information peaks appear to merge into a plateau, whereas the first two peaks corresponding to $b_{j2} = -2.0$ and $b_{j3} = 0.0$ remain unaffected. Alternatively, if I had $b_j = (-2.0, 0.0, 1.5, 1.9)$, the two right-most information peaks would merge into one, as in Figure 13.6, while, again, the first two peaks would remain relatively unaffected.

The significance of the "plateau" and merged peaks is as follows: near the ability level θ where the merger occurs, there exist two almost-simultaneous transitions between neighboring scores, i.e., from $k - 1 \rightarrow k$ and $k \rightarrow k + 1$. This suggests that around this θ, there exists a net transition from $k - 1 \rightarrow k + 1$, where subjects are not

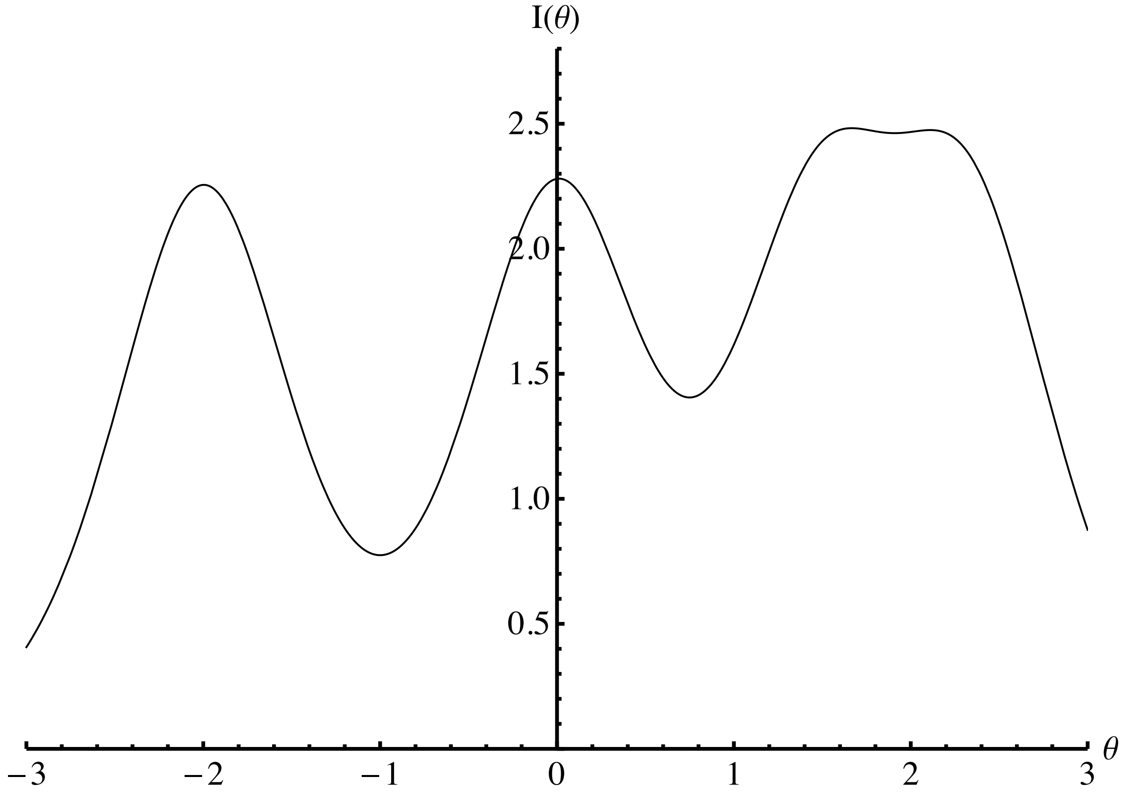


Figure 13.5. Graph of the total information for a variation of the five characteristic response curves in Figure 13.3, except with $b_j = (-2.0, 0.0, 1.5, 2.3)$

very likely to respond with score k. As I exemplify in Section 15.1.5, such a merger suggests overlapping responses between two audience groups at that ability level.

13.6 Relative difficulties of items

13.6.1 Item parameters vs. characteristic response curve intersections

In Section 13.4, I argued that I can apply the methodology of the GRM to a set of items, the sets of which I determined previously as groups of highly inter-correlated statements from the ABI. As a reminder to the reader, the three response codes (1, 2, 3) pertain to increasing degrees of misconception retainment (refer to Table 2.4). In the GRM, a typical graph of the CRCs for a single item in the set consists of three plots. Each plot represents the probability that a subject, selected arbitrarily from the total

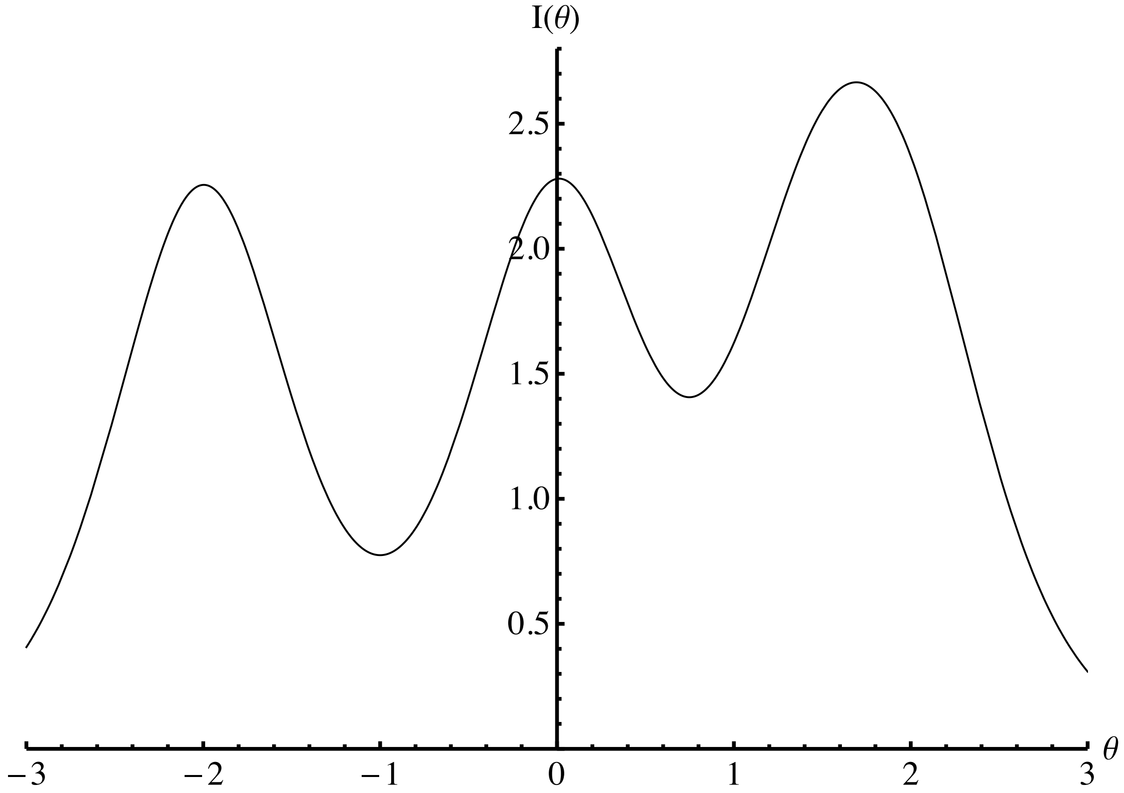


Figure 13.6. Graph of the total information for a variation of the five characteristic response curves in Figure 13.3, except with $b_{jk} = (-2.0, 0.0, 1.5, 1.9)$

sample, will respond with “1,” “2,” or “3.” The horizontal axis, generally labeled θ, represents some latent measure of the subjects. For the purposes of the ABI, that measure is “misconception endorsement.” Hence, the horizontal axis represents the tendency for the subjects to endorse the misconception associated with the item.

The item parameters provide useful measures of the locations along the misconception endorsement axis at which the probability of a response rises above 50%. For my ABI data, however, I am more interested in *transitions* between response codes. For example, I am interested in the degree of misconception endorsement, along the misconception endorsement axis, at which the probability of responding with “3,” becomes more likely than the probability of responding with “2.” More generally, I am interested in the *transition* in which the probability of responding with, e.g., score $k + 1$ becomes more likely than the probability of responding with score k.

The item parameters b_{jk} are the locations on the θ axis where the value of the second through fifth CRCs first reaches 50%. On the θ axis, the location of the *intersections* between intersecting CRCs (e.g., k and $k + 1$) appears to occur when the probability of response for each of the curves are the same as each other, but the probabilities are each *below* 50% (refer to the left-most CRC intersection, in Figure 13.5). Other responses than k or $k + 1$ may still be possible at the location of the intersections. Since the total probability of all responses at θ must sum to 100%, the probabilities of the intersecting CRCs will be below 50%, however, often only marginally so, as shown in the examples on page 234.

One fortunate consequence is that the standard error between the k^{th} and $k + 1^{\text{th}}$ CRC in the intersections is closely related to the standard error in the parameter of the $k + 1^{\text{th}}$ CRC. Consider, as an example, the plots for sA218 (see Figure 14.1). The locations on the "misconception endorsement" (or θ) axis of the item parameter for the curve labeled "2" and of the "1-2" intersection is 0.02. The percent difference in information, when evaluated at b_j vs. θ_s, is 0.1%. For the vast majority of my items, the percent difference in information, when evaluated at the location of the item parameter vs. the location of the associated CRC intersection, is much less than 1%. Hence, for the purposes of my analysis, the standard error in the intersection coordinate is essentially equivalent to the standard error in the item parameter.

13.6.2 Determination of the optimal orders of items in a group

The latent trait "misconception endorsement" represents the tendency for students to endorse the misconception associated with the item. *The tendency to retain misconceptions increases from left to right along the misconception endorsement axis.* Hence, if most of the subjects respond with "1" (dispelled the misconception prior to

AST 109) or “2” (dispelled the misconception in AST 109) for a broad range of misconception endorsement, then the item is considered relatively *easy*, because most subjects report having unlearned the associated misconception relatively early in their lives. In contrast, a relatively *hard* item is an item whose associated misconception is not abandoned readily. Specifically, if most of the subjects respond with “2” (dispelled the misconception in AST 109) or “3” (retained the misconception even after instruction in AST 109) for a broad range of misconception endorsement, then the item is considered relatively hard.

For an “easy” item, only those students who happen to have an unusually high tendency to endorse misconceptions will respond with “3.” Hence, an “easy” item can be represented graphically as three CRCs, in which the CRCs labeled “1” and “2” represent the vast majority of the students, and the CRC for “3” is well off to the right (see, for example, Figure 14.1(a), in Section 14.2.1). Likewise, for a “hard” item, only those students who happen to have an unusually low tendency to endorse misconceptions will respond with “1.” Hence, a “hard” item can be represented graphically as three CRCs, in which the CRCs labeled “2” and “3” represent the vast majority of the students, and the CRC for “1” is well off to the left (see, for example, Figure 14.2(a), in Section 14.2.3). So, the location of the “1-2” intersection for a relatively easy item will tend to occur well to the right of the location of the “1-2” intersection for a relatively hard item. Likewise, the location of the “2-3” intersection for an easy item will also tend to occur well to the right of the location of the “2-3” intersection for a relatively hard item. In summary, given a set of items, intersections between CRCs occur *farther to the right* on the misconception endorsement axis *for easier items*, and intersections between CRCs occur *farther to the left* on the misconception endorsement axis *for harder items*.

In a set of items, the locations on the “misconception endorsement” axis of the transitions produce measures of relatively easy and relatively hard items for *two* *audience groups*. Specifically, for each item in a set, the transition from "1" to "2" indicates the relative difficulty to dispel the associated misconception prior to college instruction. That is to say, each "1-2" intersection marks relative degrees of misconception endorsement in which children and adolescents are most likely to transition over from dispelling the misconception to endorsing it through one's adolescence. Likewise, for each item in a set, the transition from "2" to "3" indicates the relative difficulty to retain the misconception associated with an item, even after college instruction. That is to say, each "2-3" intersection marks relative degrees of misconception endorsement in which students in AST 109 transition over from dispelling the associated misconception in AST 109 to still retaining it even after instruction. Hence, *the "1-2" and "2-3" intersections ultimately probe misconceptions that are endorsed at two independent stages in one's life*. In a set of items, the relative sequences of the "1-2" and "2-3" intersections do not necessarily need to be the same. In other words, misconceptions that children and adolescents tend to endorse may be actually relatively common for adults to reject. Hence, *I consider two audiences separately in my analysis; the left or "1-2" intersection scores represent transitions for children and adolescents, whereas the right or "2-3" intersection scores represent transitions for students in AST 109.*

Once I determine the locations of the "1-2" and "2-3" intersections, I can proceed to sequence the items in each set by their respective locations for both categories of intersections. First, I sequence the items in the set by the locations of their "1-2" intersections. *The sequence of "1-2" intersections suggests an optimal order to teach children and adolescents about the items under consideration.* Then I sequence the items in the set by the locations of their "2-3" intersections. *The sequence of "2-3" intersections suggests an optimal order to teach adults about the items under consideration. If these two optimal orders are not the same, then the order to teach the topics is different for elementary and college-level courses.* In Section 14.2.1, I illustrate the process of starting with the coded data from the ABI to sequencing the items for a group of galaxy statements.

As outlined in Section 2.2, the response coding structure preserves a sense of timeline as to when, in one's life, one typically disabuses oneself of a misconception. The transition associated with the "1-2" intersection refers to misconceptions that are either dispelled or endorsed in one's childhood or adolescence. Likewise, the transition associated with the "2-3" intersection refers to misconceptions dispelled during or retained through instruction in AST 109. For the purposes of preserving the timeline nature of the response codes, I consider the transition associated with the "1-2" intersection to take place only for children and adolescents, and I consider the transition associated with the "2-3" intersection to take place only for adults.

13.6.3 Determination of the optimal orders of the groups themselves

Once I sequence the items in each group (e.g., in all three galaxy groups, discussed in Section 9.1), I determine the relative difficulty of the groups themselves, as a function of the audience. For both the "1-2" intersection and the "2-3" intersection, I calculate a representative weighted misconception group coordinate (MGC). The MGCs are relative measures of the *difficulty of each group* of statements, weighted by the slope parameters a_j of each of the items j within each group. As discussed in Section 13.5, statements that are the most informative also have the highest slope parameters a_j. Therefore, I choose to weight the statements by their respective slope parameters. For the "1-2" intersections, each with coordinate L, the "1-2"-intersection MGC for a group is given by

$$\mathrm{MGC}_L = \frac{\sum_j a_j L_j}{\sum_j a_j}. \tag{13.15}$$

Likewise, for the “2-3” intersections, each with coordinate R, the “2-3”-intersection MGC for a group is given by

$$\text{MGC}_R = \frac{\sum_j a_j R_j}{\sum_j a_j}. \tag{13.16}$$

The most logical order to sequence the individual groups is then in order from *highest* to *lowest* MGCs, as lower scores in each column represent harder groups in the same way that lower intersection coordinates represent harder items. As with the optimal order to sequence the items in each group, the optimal order to sequence the groups may vary as a function of the audience age.

13.7 Standard errors in the characteristic response curves

In my discussion of CRC intersections, in Section 13.5, I identified θ_s as the location on the θ axis of the intersection of two neighboring CRCs (i.e., the CRC that corresponds to score k and the CRC that corresponds to score $k + 1$). For a particular item, the *standard error* at θ_s provides a crude range outside of which θ is *significantly* above or *significantly* below θ_s. In the GRM, consider, for example, a galaxy item whose neighboring CRCs intersect at 1.4, with a standard error of 0.2. Suppose, then, that the neighboring CRCs of a second galaxy item intersect at the value 1.5, also with a standard error of 0.2. Clearly the difference between 1.4 and 1.5 is within the standard error (0.2) of each item. Therefore, neither galaxy item has a *significantly* higher θ_s than the other. On the other hand, consider, for example, two black hole items, whose respective neighboring CRC intersections take place at 1.50 and 1.85, respectively, and whose standard errors are again 0.20. One of these items is marginally significantly higher θ_s than the other, because the ranges of 1.50 ± 0.20 and 1.85 ± 0.20 just barely overlap.

In my discussion of “easy” vs. “hard” items, in Section 13.6, I explained that the locations of the CRC intersections for a set of items represents relative difficulties of easy and hard items. Correspondingly, if the locations of the CRC intersections (say, the “2-3” intersections) *between two items* lie within their standard errors, then neither one of the misconceptions associated with these items is significantly easier or harder to unlearn than the other. On the other hand, consider the locations of the CRC intersections between the two items whose intersection locations lie outside their standard errors. Then the misconception associated with the “harder” item is *significantly harder to unlearn* than the misconception associated with the “easier” item. *In essence, the sentiment that one item “feels a lot harder to learn” than another is based on whether or not the CRC intersections for these items lie within their standard errors. If they do, then the misconceptions associated with the items are of about the same difficulty. If they don’t, then one item harbors a significantly harder (or more persistent) misconception than the other.*

MULTILOG (DuToit, 2003) is software that can calculate the best fit item parameters using various models (e.g., 2PL, GRM). MULTILOG uses the maximum likelihood estimation method to estimate the reliability (or standard error, SE) of the item parameters. In Fisher information theory, it is often common practice to report the “model” standard error in the case of the maximum likelihood estimation method as

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}. \tag{13.17}$$

Since $I(\theta)$ scales as the product of $1/P(\theta)$ and the square of the slope at θ, that is, $(dP(\theta)/d\theta)^2$, the standard error scales as

$$SE(\theta) \sim \frac{1}{\sqrt{P(\theta)}} \frac{dP(\theta)}{d\theta}. \tag{13.18}$$

In the 2PL model, the SE is given by

$$SE(\theta) = \frac{1}{D\sqrt{a_j^2 P_{jk}(\theta)(1 - P_{jk}(\theta))}}, \tag{13.19}$$

where $P_{jk}(\theta)$ is the probability of responding with score k to item j, and D is a scaling factor, taken to have the value 1.7 in order to scale the estimates in the metric of the θ axis. For a comprehensive discussion of the parameter estimation algorithms performed by MULTILOG, the reader is encouraged to consult the MULTILOG manual, by duToit.

In Section 13.5, I explicitly showed, in the case of the GRM, that the location on the θ axis that maximizes the information is also *approximately* the same location at which the corresponding item parameter, b_j, is determined. Near the location of the item parameter, the probability $P_{jk}(\theta)$ is a constant approximately equal to 0.5, because each intersection occurs near the location of the item parameters, each of which are defined by where $P_{jk}(\theta) = 0.5$, as discussed in Section 13.4.1. The effect on the standard error is that the standard error at the location of the CRC intersections (θ_s) now scales as just the slope of the CRC, that is,

$$SE(\theta_s) \sim \frac{1}{2} \left. \frac{dP(\theta)}{d\theta} \right|_{\theta=\theta_s}. \tag{13.20}$$

MULTILOG, however, calculates the standard errors *exactly* at the location of the item parameters (that is, at $\theta = b_j$). One may posit that the standard errors, when evaluated at $\theta = b_j$ vs. when evaluated at $\theta = \theta_s$, may present a significant discrepancy in the quantitative features of a set of items analyzed by the model.

To examine the extent of this discrepancy, consider the 2PL model, upon which the GRM is built. For simplicity, I will rewrite the 2PL model in Equation (13.2) as

$$P(\theta) = \left(1 + e^{-a_j(\theta - b_j)}\right)^{-1}, \tag{13.21}$$

where a_j and b_j are the item parameters. Then

$$\frac{dP(\theta)}{d\theta} = \frac{a_j\, e^{-a_j(\theta-b_j)}}{\left(1 + e^{-a_j(\theta-b_j)}\right)^2}. \tag{13.22}$$

At exactly the location $\theta = b_j$,

$$\left.\frac{dP(\theta)}{d\theta}\right|_{\theta=b_j} = \frac{a_j}{4}. \tag{13.23}$$

I refer to the locations of the CRC intersections in the GRM as $\theta = \theta_s$. While the 2PL model has only a single CRC, the 2PL generalizes to the GRM by noting that each CRC maintains the same a_j, and the location on the θ axis at which neighboring CRC intersections occur is very close to the value of the corresponding item parameter. As a simplification, then, two neighboring CRCs would intersect at $\theta \approx b_j$, at which the value of $a_j(\theta - b_j)$ is small compared to 1. Performing a Taylor series expansion on Equation (13.22) gives

$$\left.\frac{dP(\theta)}{d\theta}\right|_{\theta=\theta_s} = \frac{a_j(1 - a_j(\theta - b_j))}{\left(2 - a_j(\theta - b_j)\right)^2}. \tag{13.24}$$

Factoring out 2^2 from the denominator and simplifying the result gives

$$\left.\frac{dP(\theta)}{d\theta}\right|_{\theta=\theta_s} = \frac{a_j(1 - a_j^2(\theta - b_j)^2)}{4}. \tag{13.25}$$

The discrepancy in the slope of $P(\theta)$ when θ is measured exactly at b_j vs. near b_j is given by

$$\begin{aligned}\left.\frac{dP(\theta)}{d\theta}\right|_{\theta=b_j} - \left.\frac{dP(\theta)}{d\theta}\right|_{\theta=\theta_s} &= \frac{a_j}{4}\left(1 - (1 - a_j^2(\theta - b_j)^2)\right) \\ &= \frac{a_j^3(\theta - b_j)^2}{4}. \end{aligned} \tag{13.26}$$

The discrepancy in the standard error, that is, $\delta SE = SE(\theta_s) - SE(b_j)$, is then

$$\delta SE \sim \frac{a_j^3(\theta - b_j)^2}{8}. \tag{13.27}$$

To evaluate the extent of the discrepancy, I present typical values for a_j and $\theta - b_j$ based on parameters derived from standard studies in the literature, e.g., (Cai, 2008). Table 13.1 presents the item slope parameter a_j, the difference $|\theta_s - b_j|$, the discrepancy δSE in the standard error, and the percent difference $\%\delta SE$ in the discrepancy relative to the size of the difference $|\theta_s - b_j|$.

Example	a_j	$\theta_s - b_j$	δSE	$\%\delta SE$
1	0.8	0.05	0.0003	0.16%
2	1.4	0.05	0.0017	0.49%
3	0.8	0.10	0.0013	0.64%
4	1.4	0.10	0.0069	1.97%
5	0.8	0.30	0.0155	5.84%
6	1.4	0.30	0.0617	18.43%

Table 13.1. Typical discrepancies in the reported standard errors of characteristic response curves

Table 13.1 shows that as long as (i) item discriminations (slopes) are below 1.4 and (ii) the locations on the θ axis between CRC intersections and the corresponding item parameter is less than 0.1, the standard errors at the CRC intersections vs. at the location of the item parameters are nearly the same. For these situations, one can report the standard error at $\theta = b_j$ as if the standard error was being estimated at the location of CRC intersections, without significantly affecting the validity of the reported standard error. Typical values of a_j and $\theta - b_j$ based on data from my inventory of 215 items are, respectively, $\lesssim 1.4$ and $\lesssim 0.1$.

To the interested researcher, the process by which MULTILOG calculates standard errors may be of some concern. As noted by (Cai, 2008), both MULTILOG and BILOG-MG3 are common software which can analyze data using many of the models supported by IRT. Both programs produce slightly different values of the standard errors for a number of tests; in some cases, MULTILOG gives higher standard errors than BILOG-MG3, whereas in others, MULTILOG gives lower standard errors than BILOG-MG3. Neither program is capable of producing the full covariance matrix. Discrepancies in the standard errors *between the two programs* tend to be on the order of $\lesssim 0.1$. Most interestingly, Cai (2008) notes that while multiple algorithms may be established to estimate standard errors, “an absolute gold standard is lacking in the comparison.”

Part of my dissertation project involves establishing the optimal order to teach about topics in astronomy. As I will show in the subsequent section, the methodology of IRT allows me to use the CRC intersections to establish relative measures of easy and hard items. MULTILOG is perfectly capable of calculating the item parameters quickly and efficiently. I then determine the CRC intersections by solving for the equations that represent each CRC. For a particular item, the standard error produced by MULTILOG provides a crude range outside of which θ is *significantly* above or *significantly* below θ_s.

Hence, the use of alternative software such as BILOG-MG3 seems to provide only marginal improvement in the accuracy of the standard errors.

CHAPTER 14

ITEM RESPONSE THEORY ANALYSIS OF THE GALAXY AND BLACK HOLE STATEMENT GROUPS

14.1 Overview

In Chapter 13, I presented an introduction to the methodology of IRT and argued that the GRM provides the appropriate measure with which I can analyze the probabilities of responding with various degrees of misconception retainment. In Section 13.6, I showed that the relative locations of intersections between neighboring CRCs determine measures of easy and hard items, and that the sequence of the representative transitions between degrees of misconception retainment may be unique within each set of items.

In this Chapter, I illustrate the methodology of IRT by analyzing the galaxy and black hole statement groups, which originated by performing principal components analysis as discussed in Chapter 9, using data from the Fall 2009 to Fall 2012 semesters (sample size of 401). Missing cases are excluded listwise for all subsequent analyses. Groups are analyzed in order of decreasing variance of the associated factors (refer to Section 6.4). I use MULTILOG 7 (DuToit, 2003) to (i) generate the CRCs, (ii) calculate the discrimination and item parameters (respectively, a_j and b_{jk}) on the coded responses for each item (refer to Table 2.4), and (iii) determine the standard errors in the parameters.

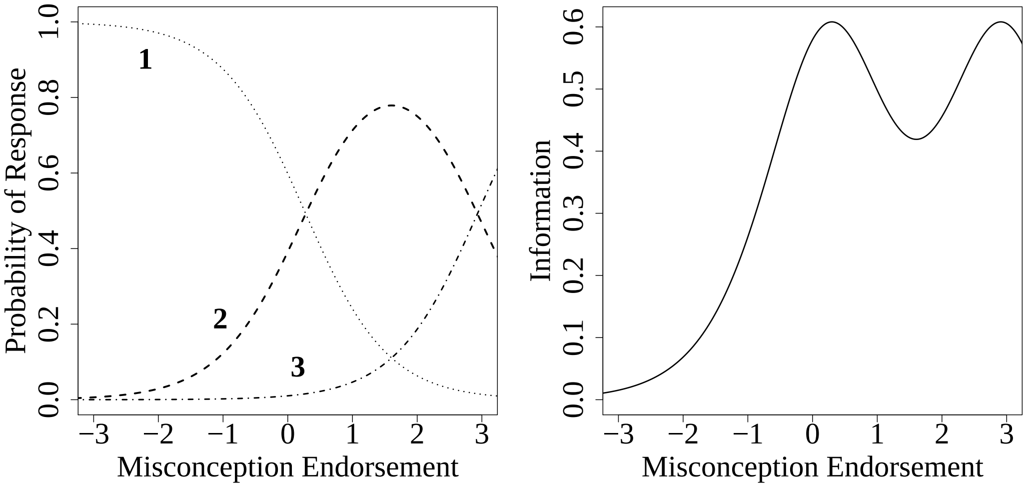
A series of plots much like those in Figure 14.1 (presented in Section 14.2.1) can be produced by determining the item parameters that best fit the GRM, then plotting the CRCs in one graph and the item information curve in a second graph. The presentation of a pair of these plots for all 215 statements of the ABI, however, would require a large amount of space in the text of this dissertation. Hence, the actual plots for all 215 items are available in Appendix L (on CD). The process with which I determine relative orders of easy and hard items is illustrated by example in Section 14.2.1.

14.2 IRT analysis of the galaxy statement groups

14.2.1 Group #1

The first group of galaxy statements consists of sA218, “the Milky Way is the only galaxy,” sA219, “the solar system is not in the Milky Way (or any other) galaxy,” sA220, “all galaxies are spiral,” sA225, “there are only a few galaxies,” sA228, “all galaxies are the same in size and shape,” sA230, “the Milky Way is just stars — no gas and dust,” and sA231, “new planets and stars don’t form today.” Plot labels “1,” “2,” and “3” respectively correspond to degrees of misconception retention, as outlined in Table 2.4. Figure 14.1 presents an example graph of CRCs, labeled by response codes 1, 2, and 3, and the item information curve, for sA218. Figure 14.1(a) contains the CRCs for various levels of misconception endorsement. Figure 14.1(b) provides a measure of how informative sA218 is for various levels of misconception endorsement. As discussed at the end of Section 13.5, the location on the “misconception endorsement” axis at which the item information curve peaks represent the approximate location of an intersection between neighboring levels of misconception retainment (i.e., retention even after instruction AST 109 vs. retention through adolescence).

As shown in Figure 14.1, there are two distinct transitions along the “misconception endorsement” axis: one for the “1-2” intersection, and one for the “2-3” intersection. The “1-2” intersection entry marks the location on the “misconception endorsement” axis where the “1-2” intersection occurs for each item in the first group. The intersection corresponds to the relative level of misconception endorsement at which children and adolescents are most likely to transition over from dispelling the misconception to endorsing it through adolescence. Likewise, the “2-3”



(a) Characteristic response curves

(b) Item information curve

Figure 14.1. Galaxy statement sA218 characteristic response curves and information curve

intersection entry marks the location on the "misconception endorsement" axis where the "2-3" intersection occurs for each item in the first group. For example, in Figure 14.1, along the "misconception endorsement" axis, the "1-2" intersection occurs at the coordinate 0.28, and the "2-3" intersection occurs at the coordinate 2.93. Each intersection corresponds to the relative level of misconception endorsement at which students in AST 109 transition from dispelling the misconception in AST 109 to still retaining it even after instruction.

Table 14.1 presents the "misconception endorsement" coordinates, along with their standard errors for the "1-2" and "2-3" intersections, using the methodology discussed in Section 13.6. For each column, an item with a *lower* coordinate is *harder* than items with higher intersection coordinates. The reported uncertainties are taken from those associated with the item parameters, whose standard errors are calculated using Equation (13.19). As discussed in Section 13.6, these standard errors provide a convenient measure of the uncertainties of the intersections themselves.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA218	1.55 ± 0.20	0.28 ± 0.10	2.93 ± 0.39
sA219	1.07 ± 0.15	−0.94 ± 0.18	2.76 ± 0.43
sA220	1.34 ± 0.17	−1.30 ± 0.18	2.56 ± 0.34
sA225	1.83 ± 0.20	−0.35 ± 0.10	2.21 ± 0.24
sA228	1.74 ± 0.20	−0.76 ± 0.12	2.63 ± 0.30
sA230	1.78 ± 0.20	−0.75 ± 0.11	2.50 ± 0.28
sA231	1.51 ± 0.17	−0.89 ± 0.14	2.56 ± 0.32

Table 14.1. Locations of the galaxy statement group #1 characteristic response curve intersections

In Section 13.6, I explained that intersections that occur farther to the right on the “misconception endorsement” axis represent easier items, and that the set of “1-2” intersections and the set of “2-3” intersections can be treated independent of each other. Table 14.1 shows that the misconception associated with sA218 is the easiest to dispel for individuals at all ages. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first galaxy group to children and adolescents, based on the “1-2” intersections, is sA218, sA225, sA228, sA230, sA231, sA219, sA220, while the order to teach to adults, based on the “2-3” intersections, is sA218, sA219, sA228, sA220, sA231, sA230, sA225.

The optimal sequences to teach the first group of galaxy statements varies as a function of the age of the audience. For example, of the seven galaxy statements in the group, sA225, “there are only a few galaxies,” is relatively easily rejected by children and adolescents, yet, of the group, sA225 is actually the most likely to be endorsed by adults, which seems rather counterintuitive. Note, however, that the “2-3” intersections for the other statements have values that are much closer to 3.0, which suggest that only those students who are the most extremely susceptible to misconceptions related to galaxies will continue to believe them even after instructions. So, for adults, sA225 may seem like an easy statement to dismiss, while the other statements are even easier to dismiss. As another example illustrating the role of audience age on relative item difficulty, sA219, “the solar system is not in the Milky Way (or any other) galaxy,” is the second most likely statement to be endorsed by children and adolescents, but is actually a relatively easy statement for adults to reject. Interestingly, sA218, “the Milky Way is the only galaxy,” is the easiest of the galaxy statements in the group for individuals of all ages to dispel.

Note that the “2-3” intersection coordinates for some statements lie within the range of their standard errors. *The standard errors for some statements are large enough that the order of some statements with very close “2-3” intersection coordinates to each other may be safely modified for instruction.* For example, the “2-3” intersection coordinates for sA220, sA230, and sA231 are all between 2.5 and 2.6, while the standard error for these statements is approximately 0.3. Hence, the exact transitions at which adults dispel the misconceptions associated with these three statements are essentially indistinguishable from each other, which is to say that the relative difficulties of unlearning the misconceptions associated with sA220, sA230, and sA231 for adults are not significantly different from each other.

This completes my analysis of the first group of galaxy statements in enough detail to demonstrate the methodology of IRT and its application in the framework of analyzing misconceptions held by college students. In the subsequent sections, I will present the results for subsequent groups more compactly. The reader is invited to refer back to Section 14.2.1 as necessary.

14.2.2 Group #2

The second group of galaxy statements consists of sA221, “the Milky Way is the center of the universe,” sA222, “the Sun is at the center of the Milky Way galaxy,” and sA224, “the Sun is at the center of the universe.” Table 14.2 presents the “misconception endorsement” coordinates, for the second group of galaxy statements, along with their standard errors for the “1-2” and “2-3” intersections. Table 14.2 shows that the three statements associated with “Being in the Center” all tend to have somewhat similar degrees of misconception endorsement to each other.

Statement	a_j	“1-2” intersection	“2-3” intersection
sA221	2.13 ± 0.22	−0.44 ± 0.09	1.58 ± 0.15
sA222	2.15 ± 0.21	−0.42 ± 0.09	1.46 ± 0.14
sA224	2.25 ± 0.23	−0.11 ± 0.08	1.65 ± 0.14

Table 14.2. Locations of the galaxy statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second galaxy statement group to children and adolescents is sA224, sA222, sA221, while the order to teach to adults is sA224, sA221, sA222. The “1-2” intersection of sA224, at −0.11, is more than three standard deviations away from the “1-2” intersections of sA221 and sA222, whose “1-2” intersections are almost indistinguishable different from each other. So, most certainly, the misconception associated with sA224 should be the first of the group to be addressed to children and adolescents. Interestingly, of the statements in the group, the “2-3” intersection coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of the statements for adults.

14.2.3 Group #3

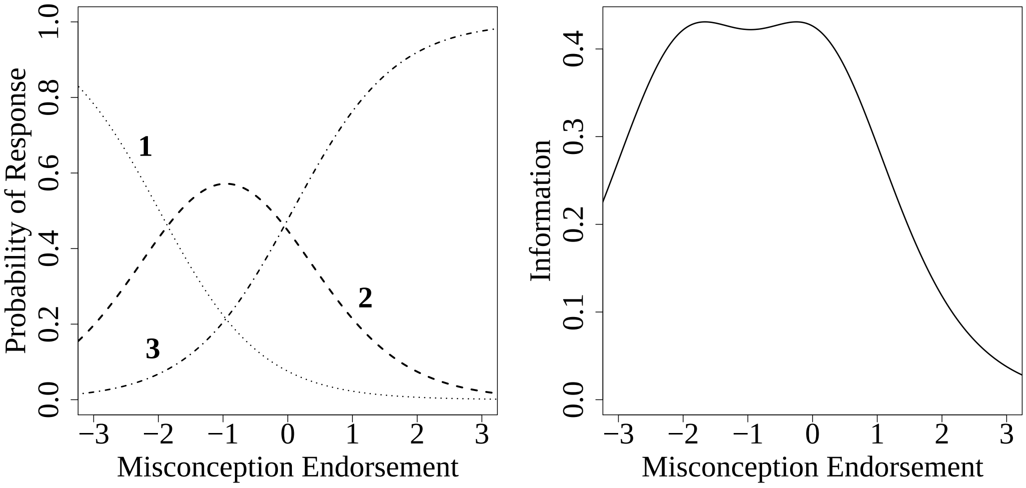
The third group of galaxy statements consists of sA226, “the galaxies are randomly distributed,” and sA227, “we see all the stars that are in the Milky Way.” Table 14.3 presents the “misconception endorsement” coordinates, for the third group of galaxy statements, along with their standard errors for the “1-2” and “2-3” intersections, as discussed in Section 13.6.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA226	1.26 ± 0.18	−1.86 ± 0.27	−0.05 ± 0.11
sA227	1.10 ± 0.17	−0.78 ± 0.18	1.93 ± 0.30

Table 14.3. Locations of the galaxy statement group #3 characteristic response curve intersections

Table 14.3 shows that the misconception associated with sA226 is much harder for adults, as well as for children and adolescents, to dispel than the misconception associated with sA227. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third galaxy statement group to both audience groups is sA227, sA226. The intersection coordinates, for either the “1-2” intersection or the “2-3” intersection, for both statements lie distinctly outside the range of their respective standard errors. Hence, there is a significant advantage to teaching sA227 prior to teaching sA226.

To illustrate that sA226 is a relatively hard item, I present a graph of the CRCs, labeled by response codes 1, 2, and 3, and the item information curve for sA226 in Figure 14.2. Consider the graph of the three CRCs, in Figure 14.2(a), where the coordinates of the “1-2” and “2-3” intersections on the “misconception endorsement” axis of sA226 are presented in Table 14.3. Compared to the locations of these same intersections for the graph in Figure 14.1(a), the intersections appear much farther to the left on the misconception endorsement axis for sA216 than they do for sA218, which indicates that sA226 is a relatively hard item, and that sA218 is a relatively easy item. These interpretations are consistent with my discussion of CRC intersections in Section 13.6. (Note, however, that no direct analytical comparisons can be made between sA218 and sA226 from the methodology of IRT, because sA218 and sA226 do not represent a locally independent set, as outlined in Section 13.3.)



(a) Characteristic response curves (b) Item information curve

Figure 14.2. Galaxy statement sA226 characteristic response curves and information curve

14.2.4 Optimal order to teach the galaxy groups

I calculate the mean MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the three galaxy groups. Table 14.4 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	−0.64	2.57
2	−0.32	1.57
3	−1.36	0.87

Table 14.4. Locations of the mean misconception group coordinates for the galaxy groups

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 1, 3, while the optimal group sequence for adults is 1, 2, 3. Essentially, the data in Table 14.4 show that of all the galaxy subtopics, those associated with the third group (sA226 and sA227), under the factor label "Spatial Distribution," should be the last to be taught to both audience groups. Interestingly, children and adolescents tend to unlearn misconceptions associated with the second group of galaxy statements, under the factor label “Being in the Center,” more readily than adults. Hence, my results suggest that the optimal order to teach galaxies, either by individual item or by group, depends on the age of the audience.

My analysis in Section 14.2 demonstrates, from the methodology of IRT, that the optimal orders to teach about galaxies is a function of the age of the audience. Table 14.5 summarizes the optimal order to teach to all 12 galaxy statements, as two separate orders: one for children and adolescents, and one for adults. For future reference, a comprehensive summary on the optimal orders to teach statements in each topic of the ABI is presented in Appendix K.

Group	Order of Statements
	Children and Adolescents
2	sA224, sA222, sA221
1	sA218, sA225, sA228, sA230, sA231, sA219, sA220
3	sA227, sA226
	Adults
1	sA218, sA219, sA228, sA220, sA231, sA230, sA225
2	sA224, sA221, sA222
3	sA227, sA226

Table 14.5. Optimal orders to teach galaxy concepts

14.3 IRT analysis of the black hole statement groups

14.3.1 Group #1

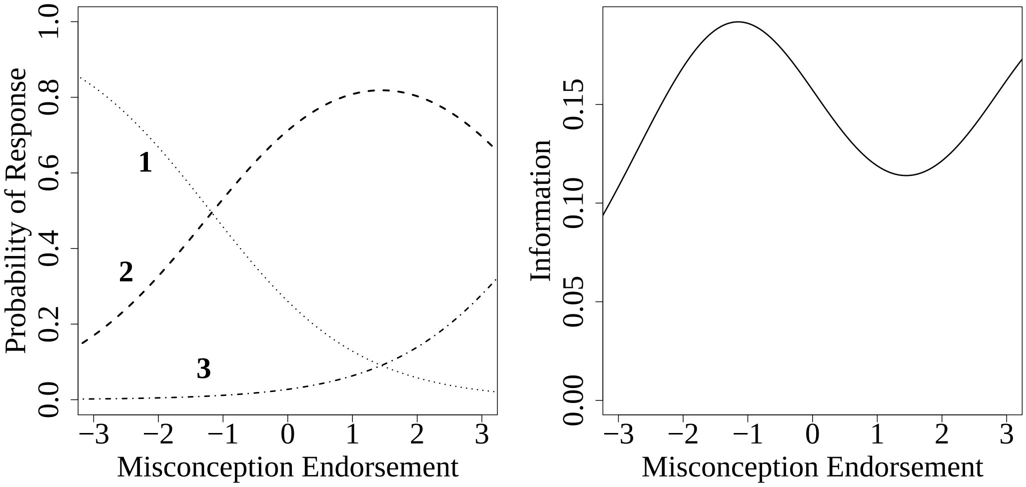
The first group of black hole statements consists of sA232, “black holes create themselves from nothing,” sA234, “black holes really don’t exist,” sA235, “black holes are empty space,” sA237, “black holes do not have mass,” and sA238, “black holes are like huge vacuum cleaners, sucking things in.” Table 14.6 presents the “misconception endorsement" coordinates for the first group of black hole statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA232	1.13 ± 0.12	-1.11 ± 0.19	2.10 ± 0.30
sA234	0.87 ± 0.14	-1.17 ± 0.26	4.07 ± 0.75
sA235	2.50 ± 0.26	-0.80 ± 0.10	0.88 ± 0.09
sA237	1.66 ± 0.19	-1.19 ± 0.14	1.24 ± 0.15
sA238	0.91 ± 0.14	-1.25 ± 0.26	0.30 ± 0.17

Table 14.6. Locations of the black hole statement group #1 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first black hole statement group to children and adolescents is sA235, sA232, sA234, sA237 sA238, while the order to teach to adults is sA234, sA232, sA237, sA235, sA238. Of the group, the misconception associated with sA238 is the hardest to dispel for both audience groups. With regard to the "1-2" intersection coordinates, Table 14.6 shows that statements sA232, sA234, sA237, and sA238 all tend to have somewhat similar degrees of misconception endorsement to each other, because their "1-2" intersection coordinates all lie within their respective standard errors. On the other hand, statement sA235 has the least negative "1-2" intersection, which indicates that the misconception associated with sA235 is the easiest of the group for children and adolescents to dispel. With regard to the "2-3" intersection coordinates, no one statement lies within the standard error of any of the others. Hence, the optimal order to teach to adults is, with certainty, sA234, sA232, sA237, sA235, sA238.

For illustration, Figure 14.3 presents the graph of CRCs, labeled by response codes 1, 2, and 3, and the item information curve for sA234. The coordinates of the "1-2" intersection and the "2-3" intersections on the "misconception endorsement" axis of sA234 are presented in Table 14.6. The discrimination parameter for sA234 $a_j = 0.87$, which, as outlined in Section 13.4.1, corresponds to an item with a low ability to discriminate between students who endorse misconceptions vs. students who reject misconceptions. The consequence is that the CRCs, in Figure 14.3(a), are less steep than the CRCs for other items. In general, the total information $I(\theta)$ also tends to be lower overall, and the uncertainty in each of the "1-2" and "2-3" intersections tends to be higher than those for other items.



(a) Characteristic response curves

(b) Item information curve

Figure 14.3. Black hole statement sA234 characteristic response curves and information curve

14.3.2 Group #2

The second group of black hole statements consists of sA233, "black holes last forever," sA242, "black holes can be seen visually, like seeing a star or planet," sA245, "black holes get bigger forever and nothing can stop them from doing so," sA246, "black holes are actual holes in space," and sA247, "a single black hole will eventually suck in all the matter in the universe." Table 14.7 presents the "misconception endorsement" coordinates for the second group of black hole statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA233	1.31 ± 0.16	−1.55 ± 0.20	1.10 ± 0.17
sA242	1.03 ± 0.17	−1.26 ± 0.22	1.98 ± 0.27
sA245	1.99 ± 0.21	−1.00 ± 0.11	1.15 ± 0.11
sA246	1.65 ± 0.17	−0.58 ± 0.11	1.37 ± 0.16
sA247	1.42 ± 0.16	−0.87 ± 0.14	1.87 ± 0.22

Table 14.7. Locations of the black hole statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second black hole statement group to children and adolescents is sA246, sA247, sA245, sA242, sA233, while the order to teach to adults is sA242, sA247, sA246, sA245, sA233. With regard to both the “1-2” intersection coordinates and the “2-3” intersection coordinates, Table 14.7 shows that both the “1-2” and the “2-3” intersection coordinates tend to lie outside of the standard errors of those statements with neighboring intersection coordinates. As examples, with regard to the “2-3” intersection coordinates, sA233 and sA245 have very similar intersection coordinates to each other, as do sA242 and sA247. No other pairs of statements in the group, however, have similar intersection coordinates.

14.3.3 Group #3

The third group of black hole statements consists of sA240, “black holes are doors to other dimensions,” sA243, “we could live in a voyage through a black hole,” and sA244, “we could travel through time in a black hole.” Table 14.8 presents the “misconception endorsement” coordinates for the third group of black hole statements, along with their standard errors for the “1-2” and “2-3” intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third black hole statement group to children and adolescents is sA240, sA244, sA243, while the order to teach to adults is sA243, sA240, sA244. These results suggest that the order to unteach incorrect ideas about the

Statement	a_j	"1-2" intersection	"2-3" intersection
sA240	1.67 ± 0.20	−0.12 ± 0.10	1.40 ± 0.16
sA243	1.22 ± 0.17	−0.69 ± 0.15	2.63 ± 0.39
sA244	2.72 ± 0.29	−0.40 ± 0.08	1.31 ± 0.11

Table 14.8. Locations of the black hole statement group #3 characteristic response curve intersections

science fiction of black holes is a function of the age of the audience. With regard to both the "1-2" intersection coordinates and the "2-3" intersection coordinates, Table 14.8 shows that the "1-2" intersection coordinates all lie outside of the standard errors of those statements with neighboring intersection coordinates. On the other hand, Table 14.8 shows that the "2-3" intersection coordinates for sA240 and sA244 lie within their respective standard errors, which suggests that the relative difficulties of unlearning the misconceptions associated with sA240 and sA244 for adults are not significantly different from each other. For adults, however, the misconception associated with sA243 is most certainly the easiest of the group to dispel. For children and adolescents, the misconception associated with sA243 is the hardest of the group to dispel.

14.3.4 Optimal order to teach the black hole groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the three black hole groups. Table 14.9 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	−1.05	1.48
2	−1.01	1.44
3	−0.38	1.62

Table 14.9. Locations of the mean misconception group coordinates for the black hole groups

In order of highest to lowest intersections, the optimal group sequence for all audience groups is 3, 2, 1, with the option to switch groups 1 and 2 for children and adolescents due to the closeness of their MGCs from the context of the standard errors of the constituent statements. Essentially, the data in Table 14.9 show that the optimal order to sequence the black hole groups does not depend on the age of the audience.

Group	Order of Statements
	Children and Adolescents
3	sA240, sA244, sA243
2	sA246, sA247, sA245, sA242, sA233
1	sA235, sA232, sA234, sA237 sA238
	Adults
3	sA243, sA240, sA244
2	sA242, sA247, sA246, sA245, sA233
1	sA234, sA232, sA237, sA235, sA238

Table 14.10. Optimal orders to teach black hole concepts

My analysis in Section 14.3 demonstrates, from the methodology of IRT, that the optimal orders to teach about black holes is a function of the age of the audience. Table 14.10 summarizes the optimal order to teach to all 13 black hole statements, as two separate orders: one for children and adolescents, and one for adults.

14.4 A brief review of the physics of black holes

In Section 14.3.4, I presented optimal orders to teach concepts related to black holes: one order for children and adolescents; and one order for adults. The orders depended marginally on the age of the audience. Of some particular interest is sA238, “black holes are like huge vacuum cleaners, sucking things in,” which, given the orders, should be the last black hole misconception to address to students in both age groups. Hence, the misconception associated with sA238 is the most persistent of all black hole misconceptions and so is the hardest to dispel. That the misconception associated with sA238 is so persistent suggests that simply explaining the concept of spacetime to students is likely insufficient for getting students to abandon the misconception entirely.

A brief physical overview of black holes is presented here, and is largely borrowed from Chapter 5 of Carroll (2003). Consider a time coordinate t and a rectangular Cartesian coordinate system, where x, y, and z are orthogonal space coordinates. Consider two frames of reference, one "unprimed" frame with coordinates (t, x, y, z), and one "primed" frame with coordinates (t', x', y', z') translating at a constant speed relative to the unprimed frame (which makes it an *inertial frame*). That four coordinates, three space and one time, are necessary to measure the separation between two points motivates the concept of *spacetime*. In a flat region of space, flat in the sense that there is no matter present to bend spacetime, the distance between any two points Δs in the unprimed frame is given by

$$(\Delta s)^2 = -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \tag{14.1}$$

where c is the speed of light in vacuum, which is constant in all reference frames. In the primed Cartesian coordinate system, the distance between two points is still Δs, because the actual separation must be independent of the frame in which one makes the measurement. The distance between any two points Δs in the primed frame is then given by

$$(\Delta s)^2 = -(c\Delta t')^2 + (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2. \tag{14.2}$$

When the separation between two points becomes infinitesimally small, $\Delta s \to ds$, so that the instantaneous displacement ds, as measured in either reference frame, is given by

$$ds^2 = -c^2\, dt^2 + dx^2 + dy^2 + dz^2. \tag{14.3}$$

The presence of matter affects the shape of spacetime, which is to say, space is stretched more in space toward wherever there is more matter. The study of how matter affects spacetime, and hence the shape of spacetime itself, is incorporated in a field known as *general relativity*.

In general relativity, one is sometimes interested in the bending of spacetime around spherical objects, of which a non-rotating black hole is well modeled (Carroll, 2003). Consider a spherical coordinate system (to be fixed later to the black hole), where r is the distance outward from the origin, ϕ is an azimuthal angle, defined in the x-y plane, and θ is the polar angle, defined by the angle downward from the z axis. In a spherically symmetric spacetime, the instantaneous displacement is given by

$$ds^2 = -c^2\, dt^2 + dr^2 + r^2\, d\theta^2 + r^2\, \sin^2\theta\, d\phi^2. \tag{14.4}$$

One then models the bending of spacetime by (i) introducing coefficients to each of the terms and (ii) introducing interaction terms, such as $dx\, dy$ and $dr\, d\phi$, with their own coefficients. In the case of a spherically symmetric gravitational field, one can avoid interacting Cartesian terms (e.g., $dx\, dy$) by using spherical coordinates, and as long as the black hole is not rotating, there are no interaction terms (e.g., $dr\, d\phi$). Consider a spherical coordinate system with its origin placed at the center of the black hole. The bending of spacetime outside of the black hole is then given by

$$ds^2 = -\left(1 - \frac{2GM}{r\,c^2}\right) dt^2 + \left(1 - \frac{2GM}{r\,c^2}\right)^{-1} dr^2 + r^2\,(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{14.5}$$

where ds is the instantaneous displacement of some particle outside of the black hole, G is the universal gravitation constant, and M is the mass contained within the black hole.

Consider the special case of a small particle that falls directly toward the black hole, so that its motion is entirely radial ($d\theta = d\phi = 0$). The instantaneous displacement of the particle is then given Carroll (2003) by

$$ds^2 = -\left(1 - \frac{2GM}{r\,c^2}\right) dt^2 + \left(1 - \frac{2GM}{r\,c^2}\right)^{-1} dr^2. \tag{14.6}$$

In the limit that the particle is far away from the black hole, $1/r \to 0$, which gives

$$ds^2 = -c^2\, dt^2 + dr^2, \tag{14.7}$$

which is the same result as that of Equation (14.4) with the angular displacements set to 0. As $r \to 2GM/c^2$, the observed displacement of the particle $ds \to 0$. That $ds \to 0$ means that the particle is never actually seen to continue to fall into the black hole, because the mass of the black hole is so great that no information from within it can escape. The quantity $2GM/c^2$ is known as the Schwarzschild radius, and is the radius inside of which particles, once fallen in, cannot reverse their trajectory and escape. The mechanism that causes particles to be unable to reverse their trajectory and escape is the extreme bending of spacetime within the black hole, due to its large mass.

Clearly, the properties of spacetime mandate that particles cannot come back out of a black hole once they have fallen into it. In a typical introductory astronomy course, professors including NFC and DJB choose not to explore the mathematics that govern the behavior of black holes, and instead characterize black holes by restating their fundamental properties in ways that are more accessible to the students. For example, professors explain the bending of spacetime around a black hole using a series of pictorial diagrams that illustrate paths taken by light. When combined with a brief lecture on the concept of spacetime, these diagrams generally provide the students with aids that adequately describe the spacetime nature of black hole.

Despite all of these aids, the misconception that “black holes are like huge vacuum cleaners, sucking things in” still persists with the students more than any other black hole misconception under consideration. Vacuum cleaners are briefly described as devices which use a pump to create a local air pressure gradient which forces particles to move toward the region of lower air pressure. The mechanisms behind vacuum cleaners and black holes are entirely different, so to a typical physicist, it seems ridiculous to make the analogy that black holes are like vacuum cleaners. Yet the observed influence by both sources on nearby particles is similar, in that particles accelerate toward the source and, once inside, do not come back out. One possibility for why the misconception persists so strongly with students is as follows: students make an attempt to understand how black holes work by establishing analogies using ordinary observations that they already understand (i.e. particles get sucked in and never come out). In other words, students tend to focus on the *observations*, rather than the *mechanism*, which is consistent with the fact that students can more easily endorse an observation, once they see it, than the mechanism, if the mechanism is less intuitive and thus not as well understood, even after instruction.

Expanding further on the notion that students focus on the observations rather than the underlying mechanism, one primary observation that students make is the tendency to expect more output from a process when the measure of input is increased. For example, one may provide some data that supports the following model: for two variables A and B, more of A should lead to more of B. Often the model is centered around one's own *observations*, and thus one forgets that there are limitations beyond which the model fails.

Models motivated by correlations between two variables can be further divided into *causal models*, in which more of A causes more of B, and *non-causal models*, in which A and B may be correlated, but one does not cause the other. Consider, for example, a causal model involving a power supply connected to an incandescent light bulb. If power input leads to the bulb becoming bright, then more power should cause the bulb to become brighter. The light bulb of course cannot get brighter indefinitely, as the filament inside the bulb would burn out at some upper limit to the power input.

Now consider, briefly, a non-causal model, such as the increase of snowshoe sales and the increase of traffic accidents. Over a period of a year, a business may find that sales of snowshoes peak during the winter, which may be when there are more traffic accidents. But the model is non-causal. For instance, if there is a surge in the sale of snowshoes in the summer, that does not mean there will be a corresponding spike in traffic accidents in the summer. Likewise, if there is a surge in the number of traffic accidents in the spring, that does not mean that customers will necessarily splurge on buying snowshoes in the spring.

As described in the preceding paragraph, both causal and non-causal models have practical limitations. In particular, limitations of a model are often forgotten if one does not invoke memory recall of the mechanism that governs the model. By virtue that one fails to recall the limitations, one often ignores the mechanism and thus has propensity to overgeneralize the model. Returning to the analogy between vacuum cleaners and black holes, because the concept of a black hole is rather abstract, one may make an analogy between a black hole and some more familiar object that shares the same function. With regard to the analogy itself, the model that one may derive from the analogy specifically is that A represents the closeness of a particle to either source, and B represents the tendency to be sucked in. As particles get closer to either source, they are more likely to be sucked in. This model clearly fails for light particles, which are not "sucked in" by a vacuum cleaner, and so students have endorsed the misconception associated with sA238 by invoking this overgeneralized model. The overgeneralization extends more significantly to the belief in the ability of a black hole to draw in matter more effectively than just due to the gravitational attraction of its mass alone, namely, that black holes have a new ability to "suck" things in, like a vacuum cleaner. Further examples of misconceptions, some of which are endorsed by students who rely on similarly overgeneralized models to explain astronomical observations, are presented in Comins (2001, 2014).

14.5 Discussion of IRT results for galaxies and black holes

In this Chapter, I illustrated the methodology of IRT by determining the optimal order to teach galaxy and black hole concepts, as a function of the age of the audience. I showed that the order to teach galaxy concepts is unique for both audience groups, with the exception of teaching concepts related to spatial distribution reasoning within or among galaxies last. I also showed that while the optimal sequence of black hole groups is the same for both audience groups, the orders of the constituent statements within each group depends on the age of the audience. I have also shown that of all the black hole misconceptions, the misconception that “black holes are like huge vacuum cleaners, sucking things in” is the hardest, because it persists the most with the students. One possible explanation for the persistence of this misconception is that students are more likely to endorse the observation that black holes “suck in” particles like vacuum cleaners than the mechanism of the bending of spacetime behind black holes, which is less intuitive.

CHAPTER 15

ITEM RESPONSE THEORY ANALYSIS OF THE OTHER STATEMENT GROUPS

In Chapter 14, I used the methodology of IRT to determine unique optimal orders to teach concepts related to galaxies and black holes. Each concept is associated with a single item. For each item, I calculated the coordinates of the intersections for adjacent characteristic response curves, where the "1-2" intersection represents a transition from rejection of a misconception before one's adolescence to endorsement of the misconception through one's adolescence, and the "2-3" intersection represents a transition from unlearning a misconception as a result of taking AST 109 to endorsing the misconception even after instruction. Misconceptions that are relatively easy to unlearn are associated with *higher* intersection coordinates, whereas misconceptions that are relatively hard to unlearn are associated with *lower* intersection coordinates. I used the intersection coordinates to show that the optimal orders for each topic depend on the age of the audience. In this Chapter, I present the results of my IRT analyses on the other sections of the ABI.

15.1 IRT analysis of the planet statement groups

15.1.1 Group #1

The first group of Planets statements (refer to Section 8.2) consists of the following five statements pertaining to misconceptions about Mars: sA161, "Mars is green (from plant life)," sA166, "Mars is the second largest planet," sA167, "life, when it did exist on Mars, was quite advanced," sA168, "there are Lowellian canals on Mars built by intelligent beings," and sA169, "Mars is Hot because it is red ... Mars — god of fire." Table 15.1 presents the "misconception endorsement" coordinates for the first group of planet statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA161	2.05 ± 0.26	−1.00 ± 0.12	2.60 ± 0.30
sA166	1.19 ± 0.14	−1.13 ± 0.18	2.16 ± 0.30
sA167	2.50 ± 0.27	−0.63 ± 0.09	1.88 ± 0.16
sA168	2.14 ± 0.26	−0.96 ± 0.11	2.61 ± 0.31
sA169	1.28 ± 0.16	−0.35 ± 0.13	3.02 ± 0.43

Table 15.1. Locations of the planets statement group #1 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Planets statement group to children and adolescents is sA169, sA167, sA168, sA161, sA166, while the order to teach to adults is sA169, sA168, sA161, sA166, sA167. With regard to the “1-2” intersections, in Table 15.1, the misconception associated with sA166 is the hardest for children and adolescents to dispel, whereas the misconception associated with sA169 is the easiest for children and adolescents to dispel. The “1-2” intersections of sA161 and sA168 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA161 and sA168 for children and adolescents are not significantly different from each other. With regard to the “2-3” intersections, in Table 15.1, the misconception associated with sA167 is the hardest for adults to reject even after instruction, whereas the misconception associated with sA169 is the easiest for adults to reject. For both audiences, sA169 represents a relatively easy item. Interestingly, sA167 is a relatively easy item for children and adolescents but a relatively hard item for adults. The “2-3” intersections of sA161 and sA168 also lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA161 and sA168 for adults are not significantly different from each other.

15.1.2 Group #2

The second group of Planets statements consists of the following three statements pertaining to misconceptions about Saturn: sA171, “Saturn is the only planet with rings,” sA172, “Saturn’s rings are solid,” and sA176, “Saturn has only one ring.” Table 15.2 presents the “misconception endorsement” coordinates for the second group of planet statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	“1-2” intersection	“2-3” intersection
sA171	1.66 ± 0.22	-0.12 ± 0.10	2.56 ± 0.31
sA172	3.50 ± 0.42	-0.18 ± 0.06	1.71 ± 0.12
sA176	1.28 ± 0.18	-0.35 ± 0.12	2.70 ± 0.38

Table 15.2. Locations of the planets statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Planets statement group to children and adolescents is sA171, sA172, sA176, while the order to teach to adults is sA176, sA171, sA172. With regard to the “1-2” intersections, Table 15.2 shows that the misconception associated with sA176 is significantly harder for children and adolescents to unlearn than the misconceptions associated with sA171 and sA172. The “1-2” intersections of sA171 and sA172 lie within their respective standard errors, indicating that neither item is significantly easier than the other for children and adolescents. With regard to the “2-3” intersections, Table 15.2 shows that the misconception associated with sA172 is significantly harder for adults to reject even after instruction than the misconceptions associated with sA171 and sA176. The “2-3” intersections of sA171 and sA176 lie within their respective standard errors, indicating that neither item is significantly easier than the other for adults.

15.1.3 Group #3

The third group of Planets statements consists of the following three statements pertaining to misconceptions about Venus: sA106, “life as we know it can exist on Venus,” sA107, “clouds on Venus are composed of water, like clouds on earth,” and sA109, “Venus is a lot like the earth in temperature.” Table 15.3 presents the “misconception endorsement” coordinates for the third group of planet statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA106	1.31 ± 0.16	-0.90 ± 0.14	2.20 ± 0.27
sA107	2.07 ± 0.23	-0.89 ± 0.10	1.47 ± 0.13
sA109	1.90 ± 0.20	-0.69 ± 0.10	1.26 ± 0.12

Table 15.3. Locations of the planets statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Planets statement group to children and adolescents is sA109, sA107, sA106, while the order to teach to adults is sA106, sA107, sA109. With regard to the “1-2” intersections, Table 15.3 shows that the misconception associated with sA109 is significantly easier for children and adolescents to unlearn than the misconceptions associated with sA106 and sA107. The “1-2” intersections of sA106 and sA107 lie within their respective standard errors, indicating that neither item is significantly harder than the other for children and adolescents. Table 15.3 also shows that the “2-3” intersections of sA107 and sA109 are marginally significantly different from each other, indicating that instructors are marginally better off teaching sA107 to adults before teaching sA109, versus the other way around. The misconception associated with sA106 is also significantly easier for adults to reject after instruction than the misconceptions associated with sA107 and sA109.

15.1.4 Group #4

The fourth group of Planets statements consists of: sA108, "Venus is very different from earth in size," sA110, "Venus is always the first star out at night," and sA170, "Mars is the sister planet to earth in physical properties and dimensions." Table 15.4 presents the "misconception endorsement" coordinates for the fourth group of planet statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA108	1.46 ± 0.18	-0.98 ± 0.15	1.01 ± 0.15
sA110	1.05 ± 0.16	-1.92 ± 0.29	1.08 ± 0.20
sA170	1.17 ± 0.16	-1.35 ± 0.22	0.66 ± 0.14

Table 15.4. Locations of the planets statement group #4 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Planets statement group to children and adolescents is sA108, sA170, sA110, while the order to teach to adults is sA110, sA108, sA170. With regard to the "1-2" intersections, Table 15.4 shows that the relative difficulties associated with each of sA108, sA110, and sA170 are significantly different from each other, as their intersection coordinates lie outside their standard errors. Hence, the optimal order to teach to children and adolescents is, with certainty, sA108, sA170, sA110. With regard to the "2-3" intersections, Table 15.4 shows that the misconception associated with sA170 is significantly harder for adults to reject even after instruction than the misconceptions associated with sA108 and sA110. The "2-3" intersections of sA108 and sA110 lie within their respective standard errors, indicating that neither item is significantly easier than the other for adults.

15.1.5 Group #5

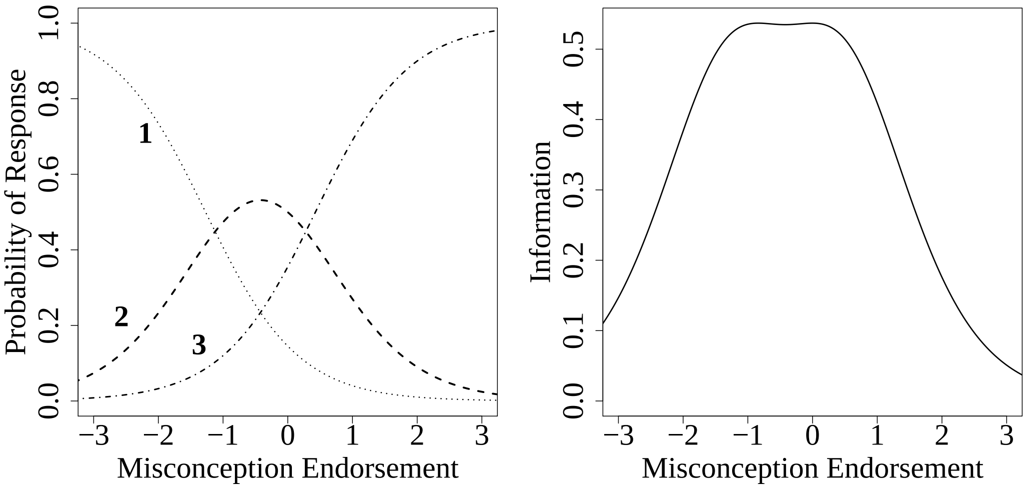
The fifth group of Planets statements consists of: sA164, “Mars has running water now,” sA165, “Mars could be made inhabitable,” and sA174, “Saturn’s rings are caused by the planet spinning so fast.” Table 15.5 presents the “misconception endorsement” coordinates for the fifth group of planet statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA164	1.42 ± 0.17	-0.93 ± 0.14	1.91 ± 0.22
sA165	1.40 ± 0.17	-1.12 ± 0.17	0.28 ± 0.12
sA174	0.84 ± 0.14	-1.44 ± 0.29	1.78 ± 0.33

Table 15.5. Locations of the planets statement group #5 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth Planets statement group to children and adolescents is sA164, sA165, sA174, while the order to teach to adults is sA164, sA175, sA165. With regard to the “1-2” intersections, Table 15.5 shows that the relative difficulties of unlearning misconceptions associated with sA165 and sA174 are marginally significantly different from each other, as are the relative difficulties associated with sA164 and sA165. Hence, the optimal order to teach to children and adolescents is, with relatively high certainty, sA164, sA165, sA174. With regard to the “2-3” intersections, Table 15.5 shows that the misconception associated with sA165 is significantly harder for adults to reject even after instruction than the misconceptions associated with sA164 and sA174. The “2-3” intersections of sA164 and sA174 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA164 and sA174 for adults are not significantly different from each other.

For illustration, Figure 15.1 presents the graph of CRCs, labeled by response codes “1,” “2,” and “3,” and the item information curve, for sA165, “Mars could be made inhabitable.” The coordinates of the “1-2” intersection and the “2-3” intersections on the “misconception endorsement” axis of sA165 are presented in Table 15.5. Note that the information curve for sA165 appears to have a plateau. The significance of the plateau, as discussed at the end of Section 13.5, is that of a net transition from response code “1” to response code “3.” if students do not realize that sA165 is false before leaving adolescence, then they are very likely to hold on to the misconception even after instruction through their adulthood. Equivalent plots for the rest of the statements in the ABI are available in Appendix L.



(a) Characteristic response curves

(b) Item information curve

Figure 15.1. Mars statement sA165 characteristic response curves and information curve

15.1.6 Optimal order to teach the planet statement groups

I calculate the MGC using the coordinates of the “1-2” intersections and the “2-3” intersections, for each of the five Planets groups. Table 15.6 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	−0.82	2.41
2	−0.20	2.12
3	−0.82	1.58
4	−1.36	0.92
5	−1.12	1.26

Table 15.6. Locations of the mean misconception group coordinates for the planet statement groups

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 1, 3, 5, 4, with the option to switch groups 1 and 3 for children and adolescents (but *not* switch groups 1 and 2), while the optimal group sequence for adults is 1, 2, 3, 5, 4. Essentially, my results suggest that the optimal order to teach about the planets, either by individual item or by group, depends on the age of the audience. Table 15.7 summarizes the optimal order to teach to all 17 Planets statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

Group	Order of Statements
	Children and Adolescents
2	sA171, sA172, sA176
1	sA169, sA167, sA168, sA161, sA166
3	sA109, sA107, sA106
5	sA164, sA165, sA174
4	sA108, sA170, sA110
	Adults
1	sA169, sA168, sA161, sA166, sA167
2	sA176, sA171, sA172
3	sA106, sA107, sA109
5	sA164, sA175, sA165
4	sA110, sA108, sA170

Table 15.7. Optimal orders to teach statements about Venus, Mars, and Saturn

15.2 IRT analysis of the star statement groups

15.2.1 Group #1

The first group of Stars statements consists of: sA20, “stars just existed — they don’t make energy or change size or color,” sA22, “all stars are stationary — fixed on the celestial sphere,” sA23, “stars emit only one color of light,” sA33, “we see the same constellations at night throughout the year,” and sA34, “stars are fixed in space.” Table 15.8 presents the “misconception endorsement” coordinates for the first group of Stars statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA20	1.69 ± 0.23	−0.21 ± 0.10	2.58 ± 0.30
sA22	2.01 ± 0.20	−0.42 ± 0.10	1.06 ± 0.11
sA23	1.28 ± 0.19	−0.37 ± 0.13	2.14 ± 0.28
sA33	1.24 ± 0.16	0.09 ± 0.13	2.30 ± 0.33
sA34	2.17 ± 0.22	0.08 ± 0.08	1.37 ± 0.13

Table 15.8. Locations of the stars statement group #1 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Stars statement group to children and adolescents is sA33, sA34, sA20, sA23, sA22, while the order to teach to adults is sA20, sA33, sA23, sA34, sA22. Table 15.8 shows that the “1-2” intersections of sA22 and sA23 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA22 and sA23 for children and adolescents are not significantly different from each other. The same can be said for sA33 and sA34. With regard to the “2-3” intersections, Table 15.8 shows that the intersection coordinates of neighboring statements in the sequence sA20, sA33, sA23, are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of these statements for children and adolescents.

15.2.2 Group #2

The second group of Stars statements consists of: sA6, “we are looking at stars as they are now,” sA7, “stars actually twinkle — change brightness,” sA11, “stars last forever,” sA17, “all stars are smaller than the Sun,” and sA18, “the galaxy, solar system and universe are the same things.” Table 15.9 presents the “misconception endorsement” coordinates for the second group of Stars statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	“1-2” intersection	“2-3” intersection
sA6	1.48 ± 0.19	0.28 ± 0.10	1.25 ± 0.17
sA7	1.04 ± 0.15	−0.84 ± 0.18	1.57 ± 0.25
sA11	1.51 ± 0.18	0.51 ± 0.11	2.76 ± 0.36
sA17	1.53 ± 0.17	0.22 ± 0.10	1.82 ± 0.21
sA18	1.10 ± 0.28	0.34 ± 0.13	3.19 ± 0.48

Table 15.9. Locations of the stars statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Stars statement group to children and adolescents is sA11, sA18, sA6, sA17, sA7, while the order to teach to adults is sA18, sA11, sA17, sA7, sA6. Table 15.9 shows that the “1-2” intersection coordinates of sA6, sA17, and sA18 lie within their respective standard errors, indicating that their relative difficulties are essentially indistinguishable from each other for children and adolescents. Also, sA7 is, by a substantial margin, the hardest item for children and adolescents, though is the second hardest item for adults.

15.2.3 Group #3

The third group of Stars statements consists of: sA3, “all of the stars are about as far away from the Earth as the Moon,” sA14, “all stars are the same distance from the Earth,” sA15, “all stars have same color and size,” sA30, “stars run on fuel:

gasoline or natural gas," and sA32, "stars follow you in your car." Table 15.10 presents the "misconception endorsement" coordinates for the third group of Stars statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA3	1.51 ± 0.19	0.02 ± 0.10	2.40 ± 0.30
sA14	2.31 ± 0.28	0.10 ± 0.08	2.68 ± 0.28
sA15	1.83 ± 0.22	0.35 ± 0.09	2.82 ± 0.36
sA30	0.87 ± 0.15	−0.82 ± 0.21	2.56 ± 0.47
sA32	1.02 ± 0.17	0.39 ± 0.15	4.98 ± 0.87

Table 15.10. Locations of the stars statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Stars statement group to children and adolescents is sA32, sA15, sA14, sA3, sA30, while the order to teach to adults is sA32, sA15, sA14, sA30, sA3. While sA32 is the easiest item for adults, the "2-3" intersection coordinates of the other four statements in the group lie within their respective standard errors, indicating that their relative difficulties are essentially indistinguishable from each other for adults.

15.2.4 Group #4

The fourth group of Stars statements consists of: sA21, "all stars end up as white dwarves," sA28, "a nova is the most powerful explosion," sA31, " 'metals' have always existed in the universe," sA35, "stars in a binary system (two stars bound together by their gravity) would quickly collide," and sA37, "all stars are isolated from all other stars (none are binary)." Table 15.11 presents the "misconception endorsement" coordinates for the fourth group of Stars statements, along with their standard errors for the "1-2" and "2-3" intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Stars statement group to children and

Statement	a_j	"1-2" intersection	"2-3" intersection
sA21	1.99 ± 0.23	−1.22 ± 0.12	1.26 ± 0.13
sA28	1.15 ± 0.15	−1.12 ± 0.15	1.46 ± 0.18
sA31	0.67 ± 0.12	−1.64 ± 0.41	0.12 ± 0.22
sA35	1.51 ± 0.19	−1.74 ± 0.18	1.36 ± 0.15
sA37	1.48 ± 0.17	−1.10 ± 0.14	2.57 ± 0.27

Table 15.11. Locations of the stars statement group #4 characteristic response curve intersections

adolescents is sA37, sA28, sA21, sA31, sA35, while the order to teach to adults is sA37, sA28, sA35, sA21, sA31. Table 15.11 shows that the "1-2" intersection coordinates of sA21, sA28, and sA37 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA21, sA28, and sA37 for children and adolescents are not significantly different from each other. With regard to the "2-3" intersections, Table 15.11 shows that the relative difficulties associated with sA21, sA28, and sA35 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA21, sA28, and sA35 for adults are not significantly different from each other.

15.2.5 Group #5

The fifth group of Stars statements consists of: sA1, "all of the stars were created at the same time," sA9, "stars have spokes," sA10, "all stars have planets," and sA13, "all stars are evenly distributed on the celestial sphere." Table 15.12 presents the "misconception endorsement" coordinates for the fifth group of Stars statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth star statement group to children and adolescents is sA1, sA9, sA10, sA13, while the order to teach to adults is sA1, sA10, sA13, sA9.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA1	0.92 ± 0.16	0.31 ± 0.15	3.25 ± 0.56
sA9	1.13 ± 0.15	−0.84 ± 0.16	1.77 ± 0.24
sA10	1.13 ± 0.16	−1.11 ± 0.18	2.16 ± 0.27
sA13	1.16 ± 0.16	−1.44 ± 0.20	1.91 ± 0.28

Table 15.12. Locations of the stars statement group #5 characteristic response curve intersections

15.2.6 Group #6

The sixth group of Stars statements consists of: sA24, “stars are closer to us than the Sun,” sA27, “all the stars in an asterism move together,” and sA29, “stars in the Milky Way are as close to each other as planets are to the Sun.” Table 15.13 presents the “misconception endorsement” coordinates for the sixth group of Stars statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA24	1.46 ± 0.19	−0.67 ± 0.13	1.92 ± 0.23
sA27	0.80 ± 0.14	−2.60 ± 0.52	−0.23 ± 0.15
sA29	2.04 ± 0.23	−0.77 ± 0.11	1.22 ± 0.13

Table 15.13. Locations of the stars statement group #6 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth star statement group to both age groups is sA24, sA29, sA27. The “1-2” intersection coordinates of sA24 and sA29 are marginally significantly different from each other, indicating that the relative difficulties of unlearning the associated misconceptions for children and adolescents are at least marginally significantly different from each other.

15.2.7 Group #7

The seventh group of Stars statements consists of: sA4, "all stars are white," sA8, "the north star is the brightest star in the sky," sA12, "the brighter a star is, the hotter it is," and sA16, "pulsars are pulsating stars." Table 15.14 presents the "misconception endorsement" coordinates for the seventh group of Stars statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA4	1.02 ± 0.17	0.46 ± 0.15	4.56 ± 0.82
sA8	1.13 ± 0.15	−0.64 ± 0.16	1.11 ± 0.20
sA12	1.23 ± 0.16	−1.42 ± 0.21	0.21 ± 0.14
sA16	0.77 ± 0.14	−2.99 ± 0.56	0.35 ± 0.20

Table 15.14. Locations of the stars statement group #7 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth star statement group to children and adolescents is sA4, sA8, sA12, sA16, while the order to teach to adults is sA4, sA8, sA16, sA12. Note however, that the "2-3" intersection coordinates of sA12 and sA16 are not significantly different from each other, indicating that the relative difficulties of unlearning the associated misconceptions for adults are at least marginally significantly different from each other. Therefore, the order to teach the topics in the seventh group of star statements only marginally depends on the age of the audience.

15.2.8 Group #8

The eighth group of Stars statements consists of: sA2, "there are 12 zodiac constellations," sA5, "the constellations are only the stars we connect to make patterns," and sA25, "there are exactly 12 constellations." Table 15.15 presents the "misconception endorsement" coordinates for the eighth group of Stars statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA2	1.87 ± 0.23	-1.25 ± 0.14	1.31 ± 0.14
sA5	0.94 ± 0.15	-1.03 ± 0.27	-0.10 ± 0.17
sA25	0.92 ± 0.16	-1.04 ± 0.23	3.54 ± 0.59

Table 15.15. Locations of the stars statement group #8 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the eighth Stars statement group to children and adolescents is sA5, sA25, sA2, while the order to teach to adults is sA25, sA2, sA5. The “1-2” intersections of sA5 and sA25 are essentially the same, indicating that the misconceptions associated with sA5 and sA25 are of about the same difficulty.

15.2.9 Optimal order to teach the star statement groups

I calculate the MGC using the coordinates of the “1-2” intersections and the “2-3” intersections, for each of the 10 Stars groups. Table 15.16 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	-0.08	1.79
2	0.15	2.09
3	0.08	2.96
4	-1.33	1.49
5	-0.83	2.22
6	-1.08	1.18
7	-1.04	1.55
8	-1.14	1.51

Table 15.16. Locations of the mean misconception group coordinates for the star statement groups

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 3, 1, 5, 7, 6, 8, 4, while the optimal group sequence for adults is 3, 5, 2, 1, 7, 8, 4, 6. Due to the closeness of their MGCs from the context of the standard errors of the constituent statements, the order of some group subsets may be re-arranged. For children and adolescents, groups 1, 2, and 3 have nearly identical MGCs, as do groups 6, 7, and 8. For adults, groups 4, 7, and 8 have nearly identical MGCs. Essentially, my results suggest that the optimal order to teach about the stars, either by individual item or by group, depends on the age of the audience. Table 15.17 summarizes the optimal order to teach to all 34 Stars statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

Group	Order of Statements
	Children and Adolescents
2	sA11, sA18, sA6, sA17, sA7
3	sA32, sA15, sA14, sA3, sA30
1	sA33, sA34, sA20, sA23, sA22
5	sA1, sA9, sA10, sA13
7	sA4, sA8, sA12, sA16
6	sA24, sA29, sA27
8	sA5, sA25, sA2
4	sA37, sA28, sA21, sA31, sA35
	Adults
3	sA32, sA15, sA14, sA30, sA3
5	sA1, sA10, sA13, sA9
2	sA18, sA11, sA17, sA7, sA6
1	sA20, sA33, sA23, sA34, sA22
7	sA4, sA8, sA16, sA12
8	sA25, sA2, sA5
4	sA37, sA28, sA35, sA21, sA31
6	sA24, sA29, sA27

Table 15.17. Optimal orders to teach statements about the stars

15.3 IRT analysis of the solar system statement groups

15.3.1 Group #1

The first group of Solar System statements consists of: sA40, “the asteroid belt is an area like we see in star wars, very densely packed,” sA41, “Mercury is so named because there is much mercury on it," sA42, "comet tails are burning — because the comet is moving so fast," sA45, "a shooting star is actually a star whizzing across the universe or falling through the sky," and sA60, "all stars have prograde rotation (spin same way as the Earth)." Table 15.18 presents the "misconception endorsement" coordinates for the first group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA40	0.99 ± 0.14	−1.42 ± 0.23	1.31 ± 0.22
sA41	1.20 ± 0.15	−0.46 ± 0.13	1.78 ± 0.24
sA42	1.73 ± 0.19	−0.58 ± 0.11	1.04 ± 0.13
sA45	1.04 ± 0.15	0.20 ± 0.14	1.03 ± 0.23
sA60	1.13 ± 0.16	−1.34 ± 0.20	2.72 ± 0.40

Table 15.18. Locations of the solar system statement group #1 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Solar System statement group to children and adolescents is sA45, sA41, sA42, sA60, sA40, while the order to teach to adults is sA60, sA41, sA40, sA42, sA45. Table 15.18 shows that the "1-2" intersections of sA40 and sA60 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA40 and sA60 for children and adolescents are not significantly different from each other. The same can almost be said for sA41 and sA42, as their "1-2" intersections are no more than marginally significantly different from each other. Table 15.18 also shows that the "2-3" intersections of sA42 and sA45 are nearly identical, indicating that the relative difficulties of unlearning the misconceptions associated with sA42 and sA45 for adults are not significantly different from each other.

15.3.2 Group #2

The second group of Solar System statements consists of: sA48, "the Solar System is the whole universe or the whole galaxy," the rather awkward statement sA56, "comets last forever," sA67, "humans have never landed a spacecraft on another planet," and sA68, "we do not have telescopes in space." Table 15.19 presents the "misconception endorsement" coordinates for the second group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA48	1.09 ± 0.16	0.30 ± 0.13	2.61 ± 0.40
sA56	1.04 ± 0.16	−0.75 ± 0.16	3.41 ± 0.52
sA67	1.28 ± 0.17	−0.08 ± 0.12	1.60 ± 0.22
sA68	2.02 ± 0.22	−0.18 ± 0.09	1.78 ± 0.17

Table 15.19. Locations of the solar system statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Solar System statement group to children and adolescents is sA48, sA67, sA68, sA56, while the order to teach to adults is sA56, sA48, sA68, sA67. Table 15.19 shows that sA56 is the hardest item of the group for children and adolescents, but is actually the easiest item for adults. With regard to both the "1-2" and "2-3" intersections, Table 15.19 shows that the intersection coordinates of sA67 and sA68 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA67 and sA68 for both age groups are not significantly different from each other.

15.3.3 Group #3

The third group of Solar System statements consists of: sA47, "the asteroid belt is between Earth and Mars," sA49, "Jupiter is almost large and massive enough to be a star," sA50, "all orbits around Sun are circular," the somewhat awkward statement sA72, “there are many galaxies in a solar system,” and sA76, “Jupiter’s great red spot is a volcano erupting.” Table 15.20 presents the “misconception endorsement” coordinates for the third group of Solar System statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	“1-2” intersection	“2-3” intersection
sA47	1.80 ± 0.20	−1.06 ± 0.13	0.95 ± 0.12
sA49	1.07 ± 0.14	−1.65 ± 0.25	0.91 ± 0.18
sA50	1.22 ± 0.16	0.12 ± 0.12	2.65 ± 0.37
sA72	1.06 ± 0.15	−1.14 ± 0.19	1.01 ± 0.19
sA76	1.09 ± 0.16	−1.01 ± 0.18	2.42 ± 0.34

Table 15.20. Locations of the solar system statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Solar System statement group to children and adolescents is sA50, sA76, sA47, sA72, sA49, while the order to teach to adults is sA50, sA76, sA72, sA49, sA47. With regard to the “2-3” intersections, Table 15.20 shows that the difficulty of the misconceptions associated with sA47, sA49, and sA72 are not significantly different from each other for adults. Therefore, the order to teach the topics in the third group of solar system statements only marginally depends on the age of the audience.

15.3.4 Group #4

The fourth group of Solar System statements consists of: sA44, “Pluto is always farther from the Sun than is Neptune,” sA46, “Jovian planets (Jupiter, Saturn, Uranus, Neptune) have solid surfaces,” sA54, “all constellations look like things they are named for,” and sA58, “Mercury (closest planet to the Sun) is hot everywhere on its surface.” Table 15.21 presents the “misconception endorsement” coordinates for the fourth group of Solar System statements, along with their standard errors for the “1-2” and “2-3” intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Solar System statement group to children and adolescents is sA54, sA46, sA44, sA58, while the order to teach to adults is sA46, sA58, sA54, sA44.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA44	1.01 ± 0.15	−0.62 ± 0.19	0.50 ± 0.18
sA46	1.42 ± 0.17	−0.51 ± 0.12	1.86 ± 0.22
sA54	0.83 ± 0.14	−0.06 ± 0.20	1.10 ± 0.29
sA58	1.46 ± 0.18	−0.81 ± 0.13	1.24 ± 0.16

Table 15.21. Locations of the solar system statement group #4 characteristic response curve intersections

15.3.5 Group #5

The fifth group of Solar System statements consists of: sA43, "there is plant life on other planets in our solar system," sA51, "planets revolve around the Earth," and sA53, "Pluto is a large, jovian (Jupiter-like) planet." Table 15.22 presents the "misconception endorsement" coordinates for the fifth group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA43	0.73 ± 0.14	−0.51 ± 0.23	2.37 ± 0.54
sA51	1.78 ± 0.23	−0.12 ± 0.09	2.46 ± 0.28
sA53	1.68 ± 0.22	−0.63 ± 0.11	2.29 ± 0.27

Table 15.22. Locations of the solar system statement group #5 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth Solar System statement group to both age groups is sA51, sA43, sA53. Therefore, the order to teach the topics in the fifth group of Solar System statements does not depend on the age of the audience. Interestingly, the "2-3" intersection coordinates of all three statements lie within their respective standard errors, so their relative difficulties are not significantly different from each other for adults.

15.3.6 Group #6

The sixth group of Solar System statements consists of: sA70, “comets are molten rock hurtling through space at high speeds and their tails are jet wash behind them,” and sA75, “comets are solid, rocky debris.” Table 15.23 presents the “misconception endorsement” coordinates for the sixth group of Solar System statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA70	1.50 ± 0.19	-1.25 ± 0.16	0.89 ± 0.14
sA75	1.46 ± 0.19	-1.15 ± 0.16	0.09 ± 0.11

Table 15.23. Locations of the solar system statement group #6 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the sixth Solar System statement group to children and adolescents is sA75, sA70, while the order to teach to adults is sA70, sA75. With regard to the “1-2” intersection coordinates, Table 15.23 shows that the intersection coordinate of sA70 is not significantly lower than that of sA75, indicating that the misconception associated with sA70 is not significantly harder for children and adolescents to dispel than that of sA75.

15.3.7 Group #7

The seventh group of Solar System statements consists of: sA62, “there are no differences between meteors, meteorites, meteoroids,” sA63, “asteroids, meteoroids, comets are same,” and sA66, “optical telescopes are the only ‘eyes’ astronomers have on the universe.” Table 15.24 presents the “misconception endorsement” coordinates for the seventh group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA62	1.75 ± 0.19	−0.55 ± 0.10	2.37 ± 0.25
sA63	1.87 ± 0.22	−0.36 ± 0.09	2.30 ± 0.25
sA66	1.00 ± 0.15	−0.47 ± 0.21	2.33 ± 0.37

Table 15.24. Locations of the solar system statement group #7 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the seventh Solar System statement group to children and adolescents is sA63, sA66, sA62, while the order to teach to adults is sA62, sA63, sA66. Note, however, that the "2-3" intersection coordinates of all three statements lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with these statements do not differ significantly from each other. Therefore, the order to teach the topics in the seventh group of Solar System statements only marginally depends on the age of the audience.

15.3.8 Group #8

The eighth group of Solar System statements consists of: sA57, "each planet has one moon," and sA59, "the day on each planet is 24 hours long." Table 15.25 presents the "misconception endorsement" coordinates for the eighth group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the eighth Solar System statement group to both age groups is sA59, sA57.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA57	1.73 ± 0.19	0.05 ± 0.10	2.27 ± 0.25
sA59	1.78 ± 0.22	0.18 ± 0.10	2.92 ± 0.25

Table 15.25. Locations of the solar system statement group #8 characteristic response curve intersections

15.3.9 Group #9

The ninth group of Solar System statements consists of: sA52, "all planets orbit exactly in the plane of the ecliptic," and sA69, "all planets have been known for hundreds of years." Table 15.26 presents the "misconception endorsement" coordinates for the ninth group of Solar System statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the ninth solar system statement group to children and adolescents is sA69, sA52, while the order to teach to adults is sA52, sA69.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA52	1.05 ± 0.16	−1.20 ± 0.21	1.60 ± 0.26
sA69	1.08 ± 0.15	−0.31 ± 0.15	1.32 ± 0.23

Table 15.26. Locations of the solar system statement group #9 characteristic response curve intersections

15.3.10 Optimal order to teach the solar system statement groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the nine Solar System groups. Table 15.27 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 8, 2, 5, 7, 4, 1, 9, 3, 6, while the optimal group sequence for adults is 1, 8, 5, 7, 2, 3, 9, 4, 6. Due to the closeness of their MGCs from the context of

Group	MGC_L	MGC_R
1	−0.70	2.96
2	−0.17	2.21
3	−0.94	1.54
4	−0.55	1.24
5	−0.35	2.37
6	−1.20	0.50
7	−0.46	2.33
8	0.12	2.60
9	−0.75	1.46

Table 15.27. Locations of the mean misconception group coordinates for the solar system statement groups

the standard errors of the constituent statements, the order of some group subsets may be re-arranged. In particular, for children and adolescents, groups 1 and 9 have nearly identical MGCs. For adults, groups 5 and 7 have nearly identical MGCs, as do groups 3 and 9. Essentially, my results suggest that the optimal order to teach about the solar system, either by individual item or by group, depends on the age of the audience. Table 15.28 summarizes the optimal order to teach to all 31 Solar System statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

15.4 IRT analysis of the moon statement groups

15.4.1 Group #1

The first group of Moon statements consists of: sA80, “craters are volcanic in origin,” sA87, “the Moon has seas and oceans of water,” sA88, “the Moon is older than the Earth: a dead planet that used to be like Earth,” sA89, “the Moon is about the same temperature as the Earth,” sA90, “the Moon has a helium atmosphere,” sA91, “the Moon has an atmosphere like the Earth,” and sA105, “the moon is lit by reflected ‘Earth light’ (that is, sunlight scattered off the Earth toward the Moon).” Table 15.29

Group	Order of Statements
	Children and Adolescents
8	sA59, sA57
2	sA48, sA67, sA68, sA56
5	sA51, sA43, sA53
7	sA63, sA66, sA62
4	sA54, sA46, sA44, sA58
1	sA45, sA41, sA42, sA60, sA40
9	sA69, sA52
3	sA50, sA76, sA47, sA72, sA49
6	sA75, sA70
	Adults
1	sA60, sA41, sA40, sA42, sA45
8	sA59, sA57
5	sA51, sA43, sA53
7	sA62, sA63, sA66
2	sA56, sA48, sA68, sA67
3	sA50, sA76, sA72, sA49, sA47
9	sA52, sA69
4	sA46, sA58, sA54, sA44
6	sA70, sA75

Table 15.28. Optimal orders to teach statements about the solar system

presents the “misconception endorsement” coordinates for the first group of Moon statements, along with their standard errors for the “1-2” and “2-3” intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Moon statement group to children and adolescents is sA91, sA89, sA87, sA80, sA88, sA105, sA90, while the order to teach to adults is sA88, sA89, sA87, sA91, sA90, sA105, sA80. Of the first five statements to teach to children and adolescents, the “1-2” intersection coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of these statements for children and adolescents. Likewise, of the first five statements to teach to adults, the “2-3” intersection

Statement	a_j	"1-2" intersection	"2-3" intersection
sA80	1.36 ± 0.16	−0.84 ± 0.15	1.62 ± 0.21
sA87	1.90 ± 0.21	−0.70 ± 0.11	2.09 ± 0.20
sA88	1.23 ± 0.15	−1.04 ± 0.17	2.30 ± 0.31
sA89	1.58 ± 0.19	−0.61 ± 0.12	2.13 ± 0.25
sA90	1.16 ± 0.14	−1.85 ± 0.25	1.99 ± 0.28
sA91	1.77 ± 0.19	−0.51 ± 0.11	2.02 ± 0.21
sA105	0.93 ± 0.13	−1.30 ± 0.25	1.63 ± 0.29

Table 15.29. Locations of the Moon statement group #1 characteristic response curve intersections

coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of these statements for adults.

15.4.2 Group #2

The second group of Moon statements consists of: sA83, "the Moon is at a fixed distance from Earth," sA85, "the Moon doesn't rotate since we see only one side of it," sA99, "a lunar month is exactly 28 days long," sA100, "at new Moon we are seeing the "far side" of the Moon," sA103, "the Moon is larger at the horizon than when it is overhead," and sA104, "the side of the moon we don't see is forever 'dark'." Table 15.30 presents the "misconception endorsement" coordinates for the second group of Moon statements, along with their standard errors for the "1-2" and "2-3" intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Moon statement group to children and adolescents is sA85, sA83, sA103, sA104, sA100, sA99, while the order to teach to adults is sA85, sA104, sA83, sA100, sA99, sA103. The "1-2" intersection coordinates of sA83 and sA85 lie within their respective standard errors, indicating that the relative

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA83	1.18 ± 0.16	−0.75 ± 0.15	1.36 ± 0.20
sA85	1.13 ± 0.16	−0.70 ± 0.16	2.21 ± 0.30
sA99	0.91 ± 0.14	−2.14 ± 0.37	−0.61 ± 0.15
sA100	1.16 ± 0.15	−1.51 ± 0.21	1.32 ± 0.19
sA103	0.76 ± 0.13	−1.07 ± 0.34	−0.67 ± 0.18
sA104	1.01 ± 0.14	−1.27 ± 0.22	1.53 ± 0.20

Table 15.30. Locations of the Moon statement group #2 characteristic response curve intersections

difficulties of unlearning the misconceptions associated with sA83 and sA85 for children and adolescents are not significantly different from each other. Likewise, the “2-3” intersection coordinates of sA83 and sA100 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA83 and sA100 for adults are not significantly different from each other. The same can be said for sA99 and sA103.

15.4.3 Group #3

The third group of Moon statements consists of: sA78, “the Moon doesn’t cause part of the tides,” sA79, “we see all sides of the Moon each month,” sA84, “the Moon changes physical shape throughout its cycle of phases,” and sA92, “the Moon has a smooth surface.” Table 15.31 presents the “misconception endorsement” coordinates for the third group of Moon statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA78	1.28 ± 0.18	−0.09 ± 0.12	2.22 ± 0.30
sA79	0.95 ± 0.15	−0.53 ± 0.16	2.64 ± 0.42
sA84	1.43 ± 0.18	−0.37 ± 0.11	2.33 ± 0.23
sA92	1.80 ± 0.21	−0.06 ± 0.09	2.47 ± 0.28

Table 15.31. Locations of the Moon statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Moon statement group to children and adolescents is sA92, sA78, sA84, sA79, while the order to teach to adults is sA79, sA92, sA84, sA78. With regard to the "1-2" intersections, Table 15.31 shows that the intersection coordinates of sA78 and sA92 lie within their respective standard errors, and that the intersection coordinates of sA79 and sA84 lie marginally within their respective standard errors. For children and adolescents, then, two subgroups, each containing two statements whose associated misconceptions are equally difficult to unlearn, can be derived from the original four Moon statements in the group. With regard to the "2-3" intersections, Table 15.31 shows that the intersection coordinates of neighboring statements in the sequence are not significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of the statements for adults.

15.4.4 Group #4

The fourth group of Moon statements consists of: sA77, "there is only one moon — ours," sA93, "the Moon sets during daylight hours and is not visible then," sA94, "there is a real man in the Moon," sA96, "because the Moon reflects sunlight, it has a mirror-like surface," and sA102, "the Moon follows you in your car." Table 15.32 presents the "misconception endorsement" coordinates for the fourth group of Moon statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA77	1.44 ± 0.18	0.65 ± 0.12	2.87 ± 0.38
sA93	1.26 ± 0.17	0.49 ± 0.12	1.85 ± 0.27
sA94	1.33 ± 0.17	0.25 ± 0.11	3.09 ± 0.42
sA96	1.12 ± 0.16	-1.14 ± 0.18	1.84 ± 0.26
sA102	1.38 ± 0.18	0.43 ± 0.11	2.83 ± 0.36

Table 15.32. Locations of the Moon statement group #4 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Moon statement group to children and adolescents is sA77, sA93, sA102, sA94, sA96, while the order to teach to adults is sA94, sA77, sA102, sA93, sA96. The "2-3" intersection coordinates of sA77, sA94, and sA102 all lie within their respective standard errors, indicating that the misconceptions associated with these statements are not significantly different from each other for adults. The same can be said about sA93 and sA96, whose "2-3" intersection coordinates are nearly identical.

15.4.5 Group #5

The fifth group of Moon statements consists of: sA97, "the Moon will someday crash into Earth," and sA98, "the Moon is a captured asteroid." Table 15.33 presents the "misconception endorsement" coordinates for the fifth group of Moon statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth Moon statement group to children and adolescents is sA98, sA97, while the order to teach to adults is sA97, sA98.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA97	2.15 ± 0.24	-1.19 ± 0.12	1.53 ± 0.13
sA98	1.91 ± 0.20	-1.04 ± 0.12	0.92 ± 0.11

Table 15.33. Locations of the Moon statement group #5 characteristic response curve intersections

15.4.6 Optimal order to teach the moon statement groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the five Moon groups. Table 15.34 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	−0.91	1.99
2	−1.21	0.99
3	−0.23	2.41
4	0.19	2.53
5	−1.12	1.24

Table 15.34. Locations of the mean misconception group coordinates for the Moon statement groups

According to Table 15.34, in order of highest to lowest intersections, the optimal group sequence for both age groups is 4, 3, 1, 5, 2. Therefore, the optimal sequence to teach the five *groups* of Moon statements is independent of the age of the audience. Due to the closeness of their MGCs from the context of the standard errors of the constituent statements, the order of some group subsets may be re-arranged. For children and adolescents, groups 2 and 5 have nearly identical MGCs. For adults, groups 3 and 4 have nearly identical MGCs. Table 15.35 summarizes the optimal order to teach to all 24 Moon statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

Group	Order of Statements
	Children and Adolescents
4	sA77, sA93, sA102, sA94, sA96
3	sA92, sA78, sA84, sA79
1	sA91, sA89, sA87, sA80, sA88, sA105, sA90
5	sA98, sA97
2	sA85, sA83, sA103, sA104, sA100, sA99
	Adults
4	sA94, sA77, sA102, sA93, sA96
3	sA79, sA92, sA84, sA78
1	sA88, sA89, sA87, sA91, sA90, sA105, sA80
5	sA97, sA98
2	sA85, sA104, sA83, sA100, sA99, sA103

Table 15.35. Optimal orders to teach statements about the Moon

15.5 IRT analysis of the earth statement groups

15.5.1 Group #1

The first group of Earth statements consists of: sA122, “X-rays can reach the ground,” sA128, “the Moon is not involved with any eclipses,” sA137, “only Earth among the planets and moons has gravity,” sA143, “the Earth will last forever,” sA152, “the Earth is the only planet with an atmosphere,” sA156, “the tides are caused just by the Earth’s rotation,” sA158, “the Sun is directly overhead everywhere on Earth at noon,” and sA159, “tides are caused just by ocean winds.” Table 15.36 presents the “misconception endorsement” coordinates for the first group of Earth statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA122	1.18 ± 0.16	−0.69 ± 0.16	1.53 ± 0.23
sA128	1.46 ± 0.17	−0.61 ± 0.12	2.61 ± 0.33
sA137	1.39 ± 0.17	−0.32 ± 0.12	1.73 ± 0.22
sA143	1.31 ± 0.17	0.48 ± 0.12	2.87 ± 0.40
sA152	1.44 ± 0.18	−0.20 ± 0.11	2.60 ± 0.34
sA156	1.59 ± 0.17	−0.63 ± 0.12	2.45 ± 0.29
sA158	0.67 ± 0.12	−0.45 ± 0.28	1.01 ± 0.38
sA159	1.59 ± 0.18	−0.41 ± 0.11	2.82 ± 0.35

Table 15.36. Locations of the Earth statement group #1 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Earth statement group to children and adolescents is sA143, sA152, sA137, sA159, sA158, sA128, sA156, sA122, while the order to teach to adults is sA143, sA159, sA128, sA152, sA156, sA137, sA122, sA158. Of the last seven statements to teach to children and adolescents, the “1-2” intersection coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of the statements for children and adolescents. Only the misconception associated with sA143 is significantly easier for children and adolescents to unlearn than misconceptions associated with other statements in the group. Likewise, of the first five statements to teach to adults, the “2-3” intersection coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconception associated with each of the statements for adults. The “2-3” intersection coordinates of sA122 and sA137 lie within their respective standard errors, indicating that the misconception associated with sA122 is not significantly harder for adults to unlearn than the misconception associated with sA137.

15.5.2 Group #2

The second group of Earth statements consists of: sA115, “Earth is at the center of the universe,” sA116, “Earth is the biggest planet,” sA133, “the sun orbits the Earth,” and sA145, “planes can fly in space.” Table 15.37 presents the “misconception endorsement” coordinates for the second group of Earth statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	“1-2” intersection	“2-3” intersection
sA115	2.76 ± 0.29	0.03 ± 0.07	2.06 ± 0.17
sA116	2.80 ± 0.28	0.05 ± 0.07	2.12 ± 0.19
sA133	2.23 ± 0.22	0.04 ± 0.09	2.09 ± 0.20
sA145	0.99 ± 0.15	0.03 ± 0.16	2.76 ± 0.50

Table 15.37. Locations of the Earth statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Earth statement group to children and adolescents is sA116, sA133, sA115, sA145, while the order to teach to adults is sA145, sA116, sA133, sA115. The “1-2” intersection coordinates of all four statements are nearly indistinguishable from each other, indicating that, for children and adolescents, the misconceptions associated with all four statements seem to be of the same difficulty. Also, the "2-3" intersection coordinates of sA115, sA116, and sA133 are all within their respective standard errors, indicating that, for adults, the difficulties of unlearning the misconception associated with each of these three statements are not significantly different from each other.

15.5.3 Group #3

The third group of Earth statements consists of: sA113, "once ozone is gone from the Earth's atmosphere, it will not be replaced," sA114, "Earth and Venus have similar atmospheres," sA126, "you can see a solar eclipse from anywhere on Earth that happens to be facing the Sun at that time," sA135, "solar eclipses happen about once a century and are seen everywhere on Earth," and sA154, "the Earth is not changing internally." Table 15.38 presents the "misconception endorsement" coordinates for the third group of Earth statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA113	1.07 ± 0.15	−1.28 ± 0.25	−0.43 ± 0.13
sA114	1.17 ± 0.14	−1.26 ± 0.20	1.08 ± 0.17
sA126	1.51 ± 0.18	−1.11 ± 0.15	0.47 ± 0.11
sA135	1.16 ± 0.15	−1.04 ± 0.17	1.68 ± 0.23
sA154	1.12 ± 0.16	−0.48 ± 0.16	1.06 ± 0.18

Table 15.38. Locations of the Earth statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Earth statement group to children and adolescents is sA154, sA135, sA126, sA114, sA113, while the order to teach to adults is sA135, sA114, sA154, sA126, sA113. The "1-2" intersection coordinates of sA113 and sA114 are nearly identical, indicating that their relative difficulties are essentially indistinguishable from each other for children and adolescents. Interestingly, sA113 is associated with the hardest misconception of the group for adults by a substantial margin. Likewise, the "2-3" intersection coordinates of sA114 and sA154 are nearly identical, indicating that their relative difficulties are essentially indistinguishable from each other for adults.

15.5.4 Group #4

The fourth group of Earth statements consists of: sA129, "the day has always been 24 hours long," sA146, "a day is exactly 24 hours long," and sA147, "a year is exactly 365 days long." Table 15.39 presents the "misconception endorsement" coordinates for the fourth group of Earth statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Earth statement group to both age groups is sA147, sA146, sA129.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA129	1.01 ± 0.15	−1.09 ± 0.22	0.36 ± 0.16
sA146	3.85 ± 0.40	−0.13 ± 0.06	0.72 ± 0.07
sA147	4.73 ± 0.53	0.01 ± 0.05	0.82 ± 0.06

Table 15.39. Locations of the Earth statement group #4 characteristic response curve intersections

15.5.5 Group #5

The fifth group of Earth statements consists of: sA141, "seasons were chosen haphazardly," sA142, "meteorites have stopped falling onto the Earth," sA148, "seasons are caused by speeding up and slowing down of Earth's rotation," sA157, "Earth has a second moon that only comes around once in awhile — 'once in a blue moon'," and sA160, "the Earth is flat." Table 15.40 presents the "misconception endorsement" coordinates for the fifth group of Earth statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA141	1.53 ± 0.18	−0.70 ± 0.12	1.71 ± 0.18
sA142	1.69 ± 0.24	−0.84 ± 0.10	2.23 ± 0.19
sA148	1.61 ± 0.18	−0.89 ± 0.13	1.83 ± 0.22
sA157	1.33 ± 0.17	−1.17 ± 0.19	3.07 ± 0.50
sA160	1.73 ± 0.17	−0.20 ± 0.19	3.23 ± 0.50

Table 15.40. Locations of the Earth statement group #5 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth Earth statement group to children and adolescents is sA160, sA141, sA142, sA148, sA157, while the order to teach to adults is sA160, sA157, sA142, sA148, sA141. The "1-2" intersection coordinates of sA142 and sA148 lie within their respective standard errors, indicating that the difficulties of unlearning the misconceptions associated with sA142 and sA148 for children and adolescents are not significantly different from each other. Likewise, the "2-3" intersection coordinates of sA157 and sA160 lie within their respective standard errors, indicating that the difficulties of unlearning the misconceptions associated with sA157 and sA160 for children and adolescents are not significantly different from each other.

15.5.6 Group #6

The sixth group of Earth statements consists of: sA118, "Spring Tide is in the spring," sA144, "the Earth's magnetic poles go through its rotation poles," and sA149, "the Earth orbits the sun at a constant speed." Table 15.41 presents the "misconception endorsement" coordinates for the sixth group of Earth statements, along with their standard errors for the "1-2" and "2-3" intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the sixth Earth statement group to both age groups is

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA118	1.23 ± 0.17	−1.08 ± 0.20	0.05 ± 0.12
sA144	1.28 ± 0.17	−1.07 ± 0.19	0.13 ± 0.12
sA149	1.66 ± 0.20	−1.22 ± 0.16	−0.30 ± 0.09

Table 15.41. Locations of the Earth statement group #6 characteristic response curve intersections

sA144, sA118, sA149. Interestingly, Table 15.41 shows that the “1-2” intersection coordinates of sA118 and sA144 are nearly indistinguishably different from each other, indicating that the misconceptions associated with sA118 and sA144 are essentially the same for children and adolescents.

15.5.7 Group #7

The seventh group of Earth statements consists of: sA125, “meteoroids enter the atmosphere a few times a night,” sA127, “auroras are caused by sunlight reflecting off polar caps,” and sA151, “the sky is blue because it reflects sunlight off oceans and lakes.” Table 15.42 presents the “misconception endorsement” coordinates for the seventh group of Earth statements, along with their standard errors for the “1-2” and “2-3” intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the seventh Earth statement group to children and adolescents is sA151, sA127, sA125, while the order to teach to adults is sA125, sA151, sA127.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA125	1.08 ± 0.16	−1.57 ± 0.25	1.17 ± 0.21
sA127	1.37 ± 0.17	−1.43 ± 0.19	0.62 ± 0.13
sA151	1.20 ± 0.17	−0.16 ± 0.13	0.93 ± 0.17

Table 15.42. Locations of the Earth statement group #7 characteristic response curve intersections

15.5.8 Group #8

The eighth group of Earth statements consists of: sA130, “the air is a blue gas,” sA131, “Halley’s comet will eventually hit Earth,” sA150, “the Earth is in the middle of the Milky Way galaxy,” and sA153, “comets affect the weather.” Table 15.43 presents the “misconception endorsement” coordinates for the eighth group of Earth statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA130	1.25 ± 0.17	−0.24 ± 0.12	2.32 ± 0.31
sA131	0.93 ± 0.15	−2.10 ± 0.34	1.60 ± 0.28
sA150	1.25 ± 0.16	−0.60 ± 0.13	2.18 ± 0.28
sA153	1.24 ± 0.16	−1.92 ± 0.23	1.82 ± 0.24

Table 15.43. Locations of the Earth statement group #8 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the eighth Earth statement group to both age groups is sA130, sA150, sA153, sA131. Table 15.43 shows that the “1-2” intersection coordinates for sA131 and sA153 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA131 and sA153 for children and adolescents are not significantly different from each other. For adults, then, two subgroups, each containing two statements whose associated misconceptions are equally difficult to unlearn, can be derived from the original four Earth statements in the group. With regard to the “2-3” intersection coordinates, Table 15.43 shows that the intersection coordinates of sA130 and sA150 lie within their respective standard errors, and that the intersection coordinates of sA131 and sA153 lie within their respective standard errors. For adults, then, two subgroups, each containing two items whose difficulties are not significantly different from each other, can be derived from the original four Earth statements in the group.

15.5.9 Group #9

The ninth group of Earth statements consists of: sA111, "Earth's axis is not tilted compared to the ecliptic," and sA112, "summer is warmer because we are closer to the sun during the summertime." Table 15.44 presents the "misconception endorsement" coordinates for the ninth group of Earth statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the ninth Earth statement group to children and adolescents is sA112, sA111, while the order to teach to adults is sA111, sA112.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA111	1.28 ± 0.18	-0.60 ± 0.13	1.52 ± 0.22
sA112	1.50 ± 0.19	-0.12 ± 0.11	0.62 ± 0.14

Table 15.44. Locations of the Earth statement group #9 characteristic response curve intersections

15.5.10 Optimal order to teach the earth statement groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the nine Earth groups. Table 15.45 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 4, 9, 1, 5, 3, 7, 6, 8, while the optimal group sequence for adults is 5, 1, 2, 8, 9, 7, 3, 4, 6. Due to the closeness of their MGCs from the context of the standard errors of the constituent statements, the order of some group subsets may be re-arranged. For children and adolescents, the following pairs of groups have very similar MGCs: 1 and 9, 3, and 7, 6 and 8. For adults, groups 1 and 5 have nearly identical MGCs, as do groups 3 and 4. Essentially, my results suggest that the optimal order to teach about the Earth, either by individual item or by group, depends on the

Group	MGC_L	MGC_R
1	−0.35	2.31
2	0.04	2.17
3	−1.04	0.77
4	−0.16	0.73
5	−0.74	2.41
6	−1.13	−0.06
7	−1.05	0.88
8	−1.15	2.00
9	−0.34	1.03

Table 15.45. Locations of the mean misconception group coordinates for the Earth statement groups

age of the audience. Table 15.46 summarizes the optimal order to teach to all 37 Earth statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

15.6 IRT analysis of the sun statement groups

15.6.1 Group #1

The first group of Sun statements consists of: sA177, "the Sun is a specific type of astronomical body with its own properties. It is not a star," sA180, "the Sun is the hottest thing in the galaxy," sA189, "the Sun is the brightest star in universe," sA190, "the Sun is the brightest object in the universe," sA198, "the Sun is the largest star," and sA209, "the Sun is hottest star." Table 15.47 presents the "misconception endorsement" coordinates for the first group of Sun statements, along with their standard errors for the "1-2" and "2-3" intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first Sun statement group to children and adolescents is sA180, sA198, sA189, sA177, sA209, sA190, while the order to teach to adults is sA177, sA209, sA189, sA198, sA190, sA180. Of the first five statements to teach to children and adolescents, the "1-2" intersection coordinates of neighboring

Group	Order of Statements
	Children and Adolescents
2	sA116, sA133, sA115, sA145
4	sA147, sA146, sA129
9	sA112, sA111
1	sA143, sA152, sA137, sA159, sA158, sA128, sA156, sA122
5	sA160, sA141, sA142, sA148, sA157
3	sA154, sA135, sA126, sA114, sA113
7	sA151, sA127, sA125
6	sA144, sA118, sA149
8	sA130, sA150, sA153, sA131
	Adults
5	sA160, sA157, sA142, sA148, sA141
1	sA143, sA159, sA128, sA152, sA156, sA137, sA122, sA158
2	sA145, sA116, sA133, sA115
8	sA130, sA150, sA153, sA131
9	sA111, sA112
7	sA125, sA151, sA127
3	sA135, sA114, sA154, sA126, sA113
4	sA147, sA146, sA129
6	sA144, sA118, sA149

Table 15.46. Optimal orders to teach statements about the Earth

statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconceptions associated with each of these statements for children and adolescents. Likewise, out of the last five statements to teach to adults, the “2-3” intersection coordinates of neighboring statements in the sequence are, at most, marginally significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconceptions associated with each of these statements for adults.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA177	1.08 ± 0.15	−0.24 ± 0.15	3.09 ± 0.45
sA180	1.91 ± 0.20	−0.12 ± 0.09	1.23 ± 0.15
sA189	2.87 ± 0.25	−0.15 ± 0.07	1.46 ± 0.12
sA190	1.77 ± 0.18	−0.48 ± 0.11	1.37 ± 0.17
sA198	3.10 ± 0.32	−0.14 ± 0.06	1.39 ± 0.11
sA209	2.55 ± 0.24	−0.34 ± 0.08	1.56 ± 0.14

Table 15.47. Locations of the Sun statement group #1 characteristic response curve intersections

15.6.2 Group #2

The second group of Sun statements consists of: sA178, "the Sun will burn forever," sA181, "the Sun does not move through space," sA182, "the Sun does not cause part of the tides," sA192, "the Sun is made of fire," and sA206, "the entire Sun is molten lava." Table 15.48 presents the "misconception endorsement" coordinates for the second group of Sun statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA178	1.60 ± 0.18	0.33 ± 0.10	2.47 ± 0.29
sA181	1.43 ± 0.17	−0.50 ± 0.13	0.60 ± 0.13
sA182	0.95 ± 0.14	−1.59 ± 0.27	1.17 ± 0.23
sA192	1.51 ± 0.20	0.54 ± 0.11	2.37 ± 0.30
sA206	1.53 ± 0.19	0.05 ± 0.10	2.51 ± 0.29

Table 15.48. Locations of the Sun statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second Sun statement group to children and adolescents is sA192, sA178, sA206, sA181, sA182, while the order to teach to adults is sA206, sA178, sA192, sA182, sA181. The "2-3" intersection coordinates of sA178, sA192, and sA206 all lie within their respective standard errors. It thus appears that the order to teach the statements in the second Sun statement group for adults may not be significantly different than the order for children and adolescents; however, both the "1-2" intersection coordinates and the "2-3" intersection coordinates of sA181 and sA182 are significantly different from each other, which implies that their respective orders for each of the age groups should remain preserved.

15.6.3 Group #3

The third group of Sun statements consists of: sA187, "Sunspots are constant fixtures on the sun," sA199, "the Sun is hottest on its surface," sA200, "the Sun has a solid core," and sA204, "the Sun's surface is perfectly uniform." Table 15.49 presents the "misconception endorsement" coordinates for the third group of Sun statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA187	1.12 ± 0.16	-1.52 ± 0.23	1.89 ± 0.27
sA199	1.64 ± 0.19	-0.88 ± 0.12	1.59 ± 0.17
sA200	1.32 ± 0.16	-1.05 ± 0.15	1.24 ± 0.14
sA204	1.22 ± 0.16	-0.82 ± 0.14	2.19 ± 0.28

Table 15.49. Locations of the Sun statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third Sun statement group to children and adolescents is sA204, sA199, sA200, sA187, while the order to teach to adults is sA204, sA187, sA199, sA200. The "1-2" intersection coordinates of sA199 and sA204 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA199 and sA204 for children and adolescents are not significantly different from each other. With regard to the "2-3" intersections, Table 15.49 shows that the relative difficulties associated with each of sA187, sA199, sA200, and sA204 are at least marginally significantly different from each other, indicating that their relative difficulties are distinctly different from each other.

15.6.4 Group #4

The fourth group of Sun statements consists of: sA193, “the Sun is a ‘heat planet’,” sA196, “the Sun is the smallest star in universe,” sA211, “it is possible that the Sun could explode in the ‘near future’,” sA214, “ the Sun is the only source of light in the galaxy — Sunlight reflects off planets and stars so we can see them,” and sA215, “Sunspots are where meteors crash into the Sun.” Table 15.50 presents the “misconception endorsement” coordinates for the fourth group of Sun statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA193	1.17 ± 0.12	−0.33 ± 0.14	2.62 ± 0.38
sA196	1.56 ± 0.20	−1.31 ± 0.16	2.74 ± 0.35
sA211	1.19 ± 0.15	−1.15 ± 0.18	1.43 ± 0.20
sA214	1.51 ± 0.18	−0.48 ± 0.12	1.54 ± 0.17
sA215	1.21 ± 0.14	−1.55 ± 0.22	3.32 ± 0.49

Table 15.50. Locations of the Sun statement group #4 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth Sun statement group to children and adolescents is sA193, sA214, sA211, sA196, sA215, while the order to teach to adults is sA215, sA196, sA193, sA214, sA211. With regard to the “1-2” intersections, Table 15.49 shows that the intersection coordinate of sA214 is marginally significantly lower than that of sA193, indicating that the misconception associated with sA214 is marginally significantly harder for children and adolescents to dispel than that of sA193. With regard to the “2-3” intersections, Table 15.49 shows that the intersection coordinates of sA193 and sA196 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA193 and sA196 for adults are not significantly different from each other. The same can be said about sA211 and sA214.

15.6.5 Group #5

The fifth group of Sun statements consists of: sA183, “sunspots are hot spots on the Sun’s surface,” sA184, “the Sun will blow up, become a black hole, and swallow the earth,” sA185, “the Sunspot cycle is 11 years long,” sA191, “the Sun always sets due west,” and sA213, “the Sun doesn’t rotate.” Table 15.51 presents the “misconception endorsement” coordinates for the fifth group of Sun statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA183	1.11 ± 0.13	−1.43 ± 0.21	1.19 ± 0.19
sA184	1.20 ± 0.15	−1.46 ± 0.20	1.80 ± 0.23
sA185	0.82 ± 0.15	−4.47 ± 0.83	0.02 ± 0.16
sA191	1.14 ± 0.16	−1.26 ± 0.23	−0.36 ± 0.12
sA213	1.29 ± 0.17	−1.04 ± 0.15	1.51 ± 0.19

Table 15.51. Locations of the Sun statement group #5 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fifth Sun statement group to children and adolescents is sA185, sA184, sA183, sA191, sA213, while the order to teach to adults is sA184, sA213, sA183, sA185, sA191. The “1-2” intersection coordinates of sA183, sA184, and sA191 are at most marginally significantly different from each other, indicating that the relative difficulties of unlearning the misconceptions associated with sA183, sA184, and sA191 for children and adolescents are not statistically significantly different from each other.

15.6.6 Group #6

The sixth group of Sun statements consists of: sA197, "the Sun has no atmosphere," sA201, "the Sun has only a few percent of the mass in the solar system," sA208, "the Sun will explode as a nova," and sA217, "it is more dangerous to look at the Sun during an eclipse because the radiation level from sun is greater then, than when there is no eclipse." Table 15.52 presents the "misconception endorsement" coordinates for the sixth group of Sun statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	**"1-2" intersection**	**"2-3" intersection**
sA197	0.90 ± 0.14	-1.83 ± 0.32	1.07 ± 0.24
sA201	1.06 ± 0.15	-1.97 ± 0.29	0.90 ± 0.17
sA208	1.29 ± 0.17	-2.05 ± 0.26	0.17 ± 0.11
sA217	1.25 ± 0.16	-1.76 ± 0.24	0.50 ± 0.13

Table 15.52. Locations of the Sun statement group #6 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the sixth Sun statement group to children and adolescents is sA217, sA197, sA201, sA208, while the order to teach to adults is sA197, sA201, sA217, sA208. With regard to the "1-2" intersections, Table 15.52 shows that the intersection coordinates of neighboring statements in the order for children and adolescence group are not significantly different from each other, indicating that there is no sudden jump in the difficulty of unlearning the misconceptions associated with each of the statements for children and adolescents. With regard to the "2-3" intersections, Table 15.52 shows that the intersection coordinates of sA197 and sA201 are marginally significantly different from each other.

15.6.7 Group #7

The seventh group of Sun statements consists of exclusively sA202, “the Sun is mostly iron.” As discussed in Section 10.6, it is possible for a single statement to load the most strongly on a particular factor. Table 15.53 presents the “misconception endorsement” coordinates for sA202, along with its standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA202	0.78 ± 0.13	−2.82 ± 0.44	3.34 ± 0.54

Table 15.53. Locations of the Sun statement group #7 characteristic response curve intersections

15.6.8 Group #8

The eighth group of Sun statements consists of: sA186, “the Sun’s surface temperature is millions of degrees Fahrenheit,” and sA188, “the Sun is yellow.” Table 15.54 presents the “misconception endorsement” coordinates for the eighth group of Sun statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA186	0.89 ± 0.16	−1.55 ± 0.36	−0.41 ± 0.14
sA188	0.69 ± 0.15	−1.23 ± 0.35	3.35 ± 0.74

Table 15.54. Locations of the Sun statement group #8 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the eighth Sun statement group to both age groups is sA188, sA186. For adults, however, the order is much more pronounced because of the higher degree of significance in the “2-3” intersection coordinates of the respective statements compared to the degree of significance in the “1-2” intersection coordinates.

15.6.9 Optimal order to teach the sun statement groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the eight Sun groups. Table 15.55 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

Group	MGC_L	MGC_R
1	-0.23	1.55
2	-0.11	1.90
3	-1.04	1.71
4	-0.96	2.32
5	-1.76	0.90
6	-1.90	0.62
7	-2.82	3.34
8	-1.42	1.23

Table 15.55. Locations of the mean misconception group coordinates for the Sun statement groups

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 1, 4, 3, 8, 5, 6, 7, while the optimal group sequence for adults is 7, 4, 2, 3, 1, 8, 5, 6. Essentially, my results suggest that the optimal order to teach about the Sun, either by individual item or by group, depends on the age of the audience. Table 15.56 summarizes the optimal order to teach to all 32 Sun statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

15.7 IRT analysis of the general astrophysics statement groups

15.7.1 Group #1

The first group of General Astrophysics statements consists of: sA255, "light travels infinitely fast," sA256, "space is infinite," sA259, "gravity is the strongest force in the universe," sA263, "astronomical ideas of mass, distance, and temperature of planets are all speculative," and sA267, "there is a center to the universe." Table 15.57

Group	Order of Statements
	Children and Adolescents
2	sA192, sA178, sA206, sA181, sA182
1	sA180, sA198, sA189, sA177, sA209, sA190
4	sA193, sA214, sA211, sA196, sA215
3	sA204, sA199, sA200, sA187
8	sA188, sA186
5	sA185, sA184, sA183, sA191, sA213
6	sA217, sA197, sA201, sA208
7	sA202
	Adults
7	sA202
4	sA215, sA196, sA193, sA214, sA211
2	sA206, sA178, sA192, sA182, sA181
3	sA204, sA187, sA199, sA200
1	sA177, sA209, sA189, sA198, sA190, sA180
8	sA188, sA186
5	sA184, sA213, sA183, sA185, sA191
6	sA197, sA201, sA217, sA208

Table 15.56. Optimal orders to teach statements about the Sun

presents the "misconception endorsement" coordinates for the first group of General Astrophysics statements, along with their standard errors for the "1-2" and "2-3" intersections.

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the first General Astrophysics statement group to children and adolescents is sA255, sA256, sA259, sA267, sA263, while the order to teach to adults is sA255, sA263, sA267, sA259, sA256. With regard to the "1-2" intersections, Table 15.57 shows that the intersection coordinates of sA263 and sA267 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA263 and sA267 for children and adolescents are not significantly different from each other. Also, the intersection coordinates for sA256 and sA259 are marginally significantly different from each

Statement	a_j	"1-2" intersection	"2-3" intersection
sA255	1.34 ± 0.15	−0.15 ± 0.12	0.68 ± 0.14
sA256	1.79 ± 0.22	−1.21 ± 0.15	−0.75 ± 0.09
sA259	1.22 ± 0.15	−1.40 ± 0.22	−0.14 ± 0.12
sA263	1.17 ± 0.16	−1.78 ± 0.25	0.53 ± 0.14
sA267	0.90 ± 0.14	−1.69 ± 0.32	0.10 ± 0.17

Table 15.57. Locations of the general astrophysics statement group #1 characteristic response curve intersections

other. With regard to the "2-3" intersections, Table 15.57 shows that the intersection coordinates of sA255 and sA263 lie within their respective standard errors, indicating that the relative difficulties of unlearning the misconceptions associated with sA255 and sA263 for adults are not significantly different from each other.

15.7.2 Group #2

The second group of General Astrophysics statements consists of: sA254, "satellites need continuous rocket power to stay in orbit around the Earth," sA258, "telescopes cannot see any details on any of the planets," sA262, "the universe as a whole is static (unchanging)," and sA270, "smaller telescopes enable astronomers to see smaller details." Table 15.58 presents the "misconception endorsement" coordinates for the second group of General Astrophysics statements, along with their standard errors for the "1-2" and "2-3" intersections.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA254	1.85 ± 0.21	−0.34 ± 0.10	1.43 ± 0.15
sA258	0.93 ± 0.15	−1.20 ± 0.22	2.98 ± 0.49
sA262	1.41 ± 0.19	−0.52 ± 0.12	2.40 ± 0.31
sA270	1.14 ± 0.17	−1.54 ± 0.25	2.57 ± 0.37

Table 15.58. Locations of the general astrophysics statement group #2 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the second General Astrophysics statement group to children and adolescents is sA254, sA262, sA258, sA270, while the order to teach to adults is sA258, sA270, sA240, sA254. With regard to the “2-3” intersection, Table 15.58 shows that the intersection coordinates of sA262 and sA270 lie within their respective standard errors, indicating that neither item is significantly easier than the other for adults.

15.7.3 Group #3

The third group of General Astrophysics statements consists of: sA252, “astronomy and astrology are the same thing,” sA271, “the most important function of a telescope is magnification,” sA272, “all space debris existing today is the result of planet collisions and explosions on planets,” and sA273, “astronomers mostly work with telescopes.” Table 15.59 presents the “misconception endorsement” coordinates for the third group of General Astrophysics statements, along with their standard errors for the “1-2” and “2-3” intersections.

Statement	a_j	**“1-2” intersection**	**“2-3” intersection**
sA252	0.82 ± 0.15	-0.04 ± 0.19	2.82 ± 0.54
sA271	1.19 ± 0.16	-1.03 ± 0.19	1.04 ± 0.18
sA272	1.42 ± 0.17	-1.52 ± 0.19	0.49 ± 0.12
sA273	1.74 ± 0.20	-1.05 ± 0.14	0.70 ± 0.11

Table 15.59. Locations of the general astrophysics statement group #3 characteristic response curve intersections

In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the third General Astrophysics statement group to both age groups is sA252, sA271, sA273, sA272. The “1-2” intersection coordinates of sA271 and sA273 lie within their respective standard errors, so the relative difficulties of unlearning the association misconceptions for adults are not significantly different from each other.

15.7.4 Group #4

The fourth group of General Astrophysics statements consists of: sA248, "cosmic rays are light rays," sA253, "gravity will eventually pull all the planets together," and sA261, "we can hear sound in space." Table 15.60 presents the "misconception endorsement" coordinates for the fourth group of General Astrophysics statements, along with their standard errors for the "1-2" and "2-3" intersections. In order of increasing tendency to endorse the associated misconception, the order to teach the statements in the fourth General Astrophysics statement group to children and adolescents is sA261, sA253, sA248, while the order to teach to adults is sA253, sA248, sA261.

Statement	a_j	"1-2" intersection	"2-3" intersection
sA248	1.08 ± 0.16	−2.11 ± 0.31	0.63 ± 0.16
sA253	1.21 ± 0.18	−1.44 ± 0.21	2.11 ± 0.28
sA261	1.00 ± 0.15	−0.91 ± 0.22	0.42 ± 0.18

Table 15.60. Locations of the general astrophysics statement group #4 characteristic response curve intersections

15.7.5 Optimal order to teach the general astrophysics statement groups

I calculate the MGC using the coordinates of the "1-2" intersections and the "2-3" intersections, for each of the four General Astrophysics groups. Table 15.61 presents the MGCs, calculated using Equations (13.15)-(13.16), for each set of intersections.

In order of highest to lowest intersections, the optimal group sequence for children and adolescents is 2, 3, 1, 4, while the optimal group sequence for adults is 2, 4, 3, 1, with the option to switch groups 3 and 4 for adults due to the closeness of their MGCs from the context of the standard errors of the constituent statements. Essentially, my results suggest that the optimal order to teach about astrophysics in

Group	$\boldsymbol{MGC_L}$	$\boldsymbol{MGC_R}$
1	−1.20	0.02
2	−0.80	2.20
3	−1.02	1.06
4	−1.50	1.11

Table 15.61. Locations of the mean misconception group coordinates for the general astrophysics statement groups

general, either by individual item or by group, depends on the age of the audience. Table 15.62 summarizes the optimal order to teach to all 16 General Astrophysics statements under consideration, as two separate orders: one for children and adolescents, and one for adults.

Group	Order of Statements
	Children and Adolescents
2	sA254, sA262, sA258, sA270
3	sA252, sA271, sA273, sA272
1	sA255, sA256, sA259, sA267, sA263
4	sA261, sA253, sA248
	Adults
2	sA258, sA270, sA240, sA254
4	sA253, sA248, sA261
3	sA252, sA271, sA273, sA272
1	sA255, sA263, sA267, sA259, sA256

Table 15.62. Optimal orders to teach statements about astrophysics in general

CHAPTER 16

PARTIAL CROSSOVER TEST OF THE ITEM RESPONSE THEORY RESULTS ON BLACK HOLE AND GALAXY INSTRUCTION

In Chapter 14, I presented optimal orders to teach statements about galaxies and black holes using the methodology of item response theory. In this Chapter, I present the design of a partial crossover analysis, using video lectures to teach about galaxies and black holes, along with associated pretests and posttests tailored to each video lecture. Then I show the effect of instruction re-arrangement on how much students scores improve relative to the pretest scores.

16.1 Design of the partial crossover test

16.1.1 Test and video lecture preparation

After I collected data from the ABI during the Fall 2009 to Fall 2011 semesters, I proposed tests of the optimal orders to teach about black holes and galaxies, which I administered to the students taking AST 109 in the Spring 2012 semester. As discussed in Chapter 4, test questions on black holes and galaxies were designed, revised, and then implemented in the BHGP, with the questions and response options specifically tailored to the statements on the inventory and incorrect beliefs held by the students. I used essentially these questions to design a multiple-choice pretest prior to the video lecture based on these questions, then I designed a posttest in which the questions and response options for both topics were randomized compared to the pretest. The pretest questions and response options for both black holes and galaxies are presented in Appendix J. Between the video lecture tests and the BHGP, only the last two galaxy questions are slightly different. All other questions remain the same.

On the day of a particular video lecture, the questions for the pretest were presented to the class one question at a time. All students were first presented with a pretest on the accompanying topic immediately before the video lecture and then with the corresponding posttest immediately following the video lecture, all in a single class. Students responded to each question by indicating the letter corresponding to one of the response options as their answer. For example, if, on Question 1, a student believed that B was the correct answer, the student wrote the letter B. Students were not allowed to consult any references during any of the tests. No extra credit was awarded to the students. Students were told that right or wrong answers would not impact their grade in any way, and that by participating, students would earn attendance credit (which is mandatory in the class and not "extra" credit) in the class for the day.

A video lecture for each topic came between the administrations of the associated pretests and posttests. Volunteering himself, DJB prepared video lectures, about 50-60 minutes long, for each topic. The video lectures were presented to the students in place of traditional live lecture for instruction on black holes and galaxies. The instruction in the videos covered all of the black hole and galaxy statements under consideration in the ABI. Two videos were produced for each topic, with the only difference in the videos for each topic being the *arrangement of instruction*. For each topic, a video was produced where DJB arranged the instruction of each topic in an order that best represented his conventional live instruction. The resulting video represents the *conventional* order. A second video was then produced where I suggested specific re-arrangements of the instruction within the video based on the results of my item response theory tests, as presented in Chapter 14. The resulting video represents the *alternate* order. DJB was not told that the re-arrangements were the result of my item response theory analyses. No other changes to the videos were made for each topic.

At the time of designing these questions, the optimal order to teach black holes and galaxies was based on three semesters worth of data. Because I had not yet included a fourth semester of data, I determined the optimal order to teach black holes and galaxies based on the first three semesters of data, using the same procedure that I outlined in Section 13.6. I have since examined the extent to which the inclusion of the Fall 2012 data would have affected the factor structure produced by PCA on the Fall 2009 to Fall 2011 data, coded in the manner described in Table 2.4. I calculated the difference in correlations of the galaxies and black hole statements using the procedure discussed in Chapter 7, with the Fall 2009 to Fall 2011 responses representing one semester group and the Fall 2009 to Fall 2012 responses representing the other semester group. Between the semester groups, the mean difference in galaxy statement correlations is $M_s = 0.035$, and the mean difference in black hole statement correlations is $M_s = 0.024$, indicating that the inclusion of the Fall 2012 semester data affects the magnitude of statement correlations only by about 0.03.

The calculated optimal orders to teach black hole and galaxy statements based on the two semester sets, however, were different in some ways. For example, PCA produced four black hole factors using data from Fall 2009 to Fall 2011 semesters, instead of three black hole factors using data from all four semesters. Fortunately, data from both semester sets produced the same groupings of galaxy statements. In other words, between the two semester sets, no galaxy statements were exchanged (swapped) among the groups. The loadings of the statements within each group changed, only marginally (typically within 0.1). These changes were so small that the sequence of galaxy statements within a group, from highest to lowest factor loading, changed only marginally. Within all groups between the semester sets, the relative difficulties of the black hole and galaxy items were not significantly affected, in that the characteristic response curve intersection coordinates (refer to Section 13.6) of any re-arranged statements were within their standard errors. Therefore, I consider the sequence of the alternate video lecture instruction to be a reasonably close representation of the sequence that I would have suggested if I had included data from the Fall 2012 semester.

16.1.2 Test design modification

In the Spring 2012 semester, black holes were taught before galaxies. Originally I had planned on a crossover study, in which the class was divided into two nearly equal halves, selected partly at random by alternating first letter of the student's last name. Each half was assigned to one of two identical classrooms in the same building, with both rooms being smaller than that of their customary lecture classroom. The students in the rooms were to receive either the *conventional* or *alternate* video lecture on black holes. Then the students who received the conventional black holes video lecture would receive the alternate galaxies video lecture, and the students who received the alternate black holes video lecture would receive the conventional galaxies video lecture.

Due to a miscommunication on the day of the administration of the black holes video lectures, the alternate black hole video was shown to students in both lecture rooms. So, I saved the situation by changing the design of the study from a full crossover analysis to a partial crossover analysis, as follows: In the Spring 2012 semester, all students received the *alternate* black holes video lecture, in the separate rooms, and the *alternate* galaxies video lecture, together in the same room. In the Spring 2013 semester, again taught by DJB, all students received the video lectures together in the same room. Students first received the *alternate* black holes video lecture, but then all students received the *conventional* galaxies video lecture. The partial crossover study benefits from having a larger total sample size and from using the same alternate black holes video lecture as a cross-reference. Table 16.1

summarizes the actual design of the partial crossover study and lists the number of students whose data were used for *both topics*, with respect to each semester.

Semester	Sample Size	Topic	Instruction Sequence
Spring 2012	114	Black Holes Galaxies	Alternate Alternate
Spring 2013	94	Black Holes Galaxies	Alternate Conventional

Table 16.1. Partial crossover study design of instruction on black holes and galaxies

16.2 Performance on individual questions

16.2.1 Black holes

In both the Spring 2012 and Spring 2013 semesters, students in AST 109 received the alternate black holes video lecture. For each black holes pretest and posttest question (refer to Appendix J), I calculated the mean fraction of students in the class who answered each question correctly. Then I compared the fractions correct for each question between the two semesters to test for significant differences. Table 16.2 presents the mean fractions correct for each question individually, as well as the change (mean of posttest minus pretest scores), rounded off, and their standard deviations.

Note that the subtopic associated with Question 5, that black holes do not suck, which appears at the end of the alternate video lecture, was not presented to the students in the Spring 2013 semester, because we ran out of time, and so we cut the video lecture short immediately prior to its instruction. The last subtopic, which conveys the idea that black holes do not suck, was altogether skipped. Hence, in Table 16.2, there was relatively little change in test scores on Question 5 in the Spring 2013 semester. To quantify the significance of the gain for each question between the semesters, I performed an ANOVA test on the gains, one question at a time. The

Question	Test	Spring 2012 n = 144		Spring 2013 n = 133	
		M	***SD***	***M***	***SD***
1. How are black holes created today?	Pretest	0.486	0.502	0.361	0.482
	Posttest	0.639	0.482	0.496	0.502
	Change	0.153	0.571	0.135	0.588
2. What is the fate of black holes?	Pretest	0.243	0.430	0.226	0.420
	Posttest	0.931	0.255	0.910	0.288
	Change	0.688	0.480	0.684	0.513
3. Black holes consist of:	Pretest	0.778	0.417	0.692	0.464
	Posttest	0.819	0.386	0.707	0.457
	Change	0.042	0.486	0.015	0.603
4. As detectable from our universe, what shape does a black hole have?	Pretest	0.056	0.230	0.135	0.343
	Posttest	0.160	0.368	0.158	0.366
	Change	0.104	0.405	0.023	0.484
5. A spacecraft is put into circular orbit around a distant black hole. Which one of the following outcomes will happen to the spacecraft?	Pretest	0.076	0.267	0.075	0.265
	Posttest	0.345	0.477	0.113	0.318
	Change	0.269	0.518	0.038	0.258
6. What would happen to an asteroid that passed into a black hole?	Pretest	0.799	0.402	0.805	0.398
	Posttest	0.840	0.368	0.872	0.335
	Change	0.042	0.471	0.068	0.464
7. How do astronomers detect black holes?	Pretest	0.847	0.361	0.827	0.380
	Posttest	0.769	0.423	0.857	0.351
	Change	−0.078	0.505	0.030	0.443
8. If we boarded a spaceship to journey into a black hole, what would happen to us?	Pretest	0.806	0.397	0.789	0.409
	Posttest	0.804	0.398	0.812	0.392
	Change	−0.001	0.517	0.023	0.529

Table 16.2. Mean fraction of correct pretest and posttest answers associated with the alternate black holes video lecture

results are presented in Table 16.3, where df_2 is the within-subjects degrees of freedom, and η^2 and ω^2 are estimations of the effect size, as discussed in Section 3.2.

The data in Table 16.2 and Table 16.3 show that for seven of the eight questions, there were no significant differences in posttest scores between the Spring 2012 and Spring 2013 semesters. Table 16.3 also shows that Question 5 appears to harbor the only significant score change ($p < 0.0005$). Not surprisingly, the results suggest that even pre-recorded instruction is significantly better than no instruction on the topic at all. The difference between pretest and posttest scores for Question 7, between the two semesters, is marginally different. The reason for the nearly significant difference is unknown. No other questions had significant semester differences in score changes.

Recall that students in the Spring 2012 semester were selected partly at random to watch the black holes video lecture in one of two identical rooms, neither of which was their customary lecture classroom. In the Spring 2013 semester, all students received the same black holes video, again in the alternate sequence, but together as a class in their normal lecture classroom. Based on my results, one should not expect any significant difference in student performance when students receive instruction either in their customary classroom or in an alternate classroom suitable for the same lecture.

16.2.2 Galaxies

In the Spring 2012 semester, students in AST 109 received the alternate galaxies video lecture, whereas in the Spring 2013 semester, students received the conventional galaxies video lecture. There was adequate class time to present each video in its entirety to the students in both semesters. For each galaxies pretest and posttest question (refer to Appendix J), I calculated the mean fraction of students in the class who answered each question correctly. Then I compared the fractions correct for

Question	Spring 2012 vs. Spring 2013				
	df_2	F	p	η^2	ω^2
1. How are black holes created today?	275	0.063	0.802	0.000	−0.003
2. What is the fate of black holes?	275	0.003	0.956	0.000	−0.004
3. Black holes consist of:	275	0.165	0.685	0.001	−0.003
4. As detectable from our universe, what shape does a black hole have?	275	2.329	0.128	0.008	0.005
5. A spacecraft is put into circular orbit around a distant black hole. Which one of the following outcomes will happen to the spacecraft?	273	21.266	<.0005**	0.072	0.069
6. What would happen to an asteroid that passed into a black hole?	275	0.214	0.644	0.001	−0.003
7. How do astronomers detect black holes?	274	3.485	0.063	0.013	0.001
8. If we boarded a spaceship to journey into a black hole, what would happen to us?	274	0.128	0.721	0.000	−0.003

Table 16.3. Tests of significant gains associated with the alternate black holes video lecture

each question between the two semesters to test for significant differences. Table 16.4 presents the mean fractions correct for each question individually, as well as the change (mean of posttest minus pretest scores), rounded off, and their standard deviations.

As a reminder to the reader, the alternate galaxies video lecture was presented in the Spring 2012 semester, while the conventional galaxies video lecture was presented in the Spring 2013 semester. To quantify the significance of the gain for each question between the semesters, I performed an ANOVA test on the gains, one question at a time. The results are presented in Table 16.5, where again df_2 is the within-subjects degrees of freedom, and η^2 and ω^2 are estimations of the effect size.

Table 16.4 and Table 16.5 show that, for the most part, the re-arrangement of subtopics within the galaxies video lecture appears to have no significant effects on student performance. Only Questions 3 and 4 harbor *at least* marginally significant differences in student performance. No other questions had significant semester differences in score changes. In Question 3, students who watched the conventional galaxies video lecture performed *significantly worse* than students who watched the alternate galaxies video lecture. Interestingly, however, with regard to Question 4, the performance of students who watched the conventional video lecture was marginally significantly better than the performance of students who watched the alternate video lecture. (Because one of the galaxy video lectures has been lost, I cannot present the timestamps for when this material was covered between the two galaxy videos.) Therefore, the re-arrangement of these questions may have caused this “switcharoo” effect on the score changes between the two questions.

		Spring 2012 *n* = 138		Spring 2013 *n* = 109	
Question	**Test**	***M***	***SD***	***M***	***SD***
1. How many galaxies exist in the universe today?	Pretest	0.741	0.440	0.743	0.439
	Posttest	0.862	0.346	0.917	0.277
	Change	0.121	0.505	0.174	0.488
2. Relative to the Milky Way galaxy, the solar system is located:	Pretest	0.273	0.447	0.165	0.373
	Posttest	0.514	0.502	0.459	0.501
	Change	0.241	0.668	0.294	0.582
3. How many distinct shapes do galaxies have?	Pretest	0.827	0.379	0.872	0.336
	Posttest	0.826	0.380	0.716	0.453
	Change	−0.001	0.452	−0.156	0.530
4. Where in the universe is the Milky Way located today?	Pretest	0.719	0.451	0.697	0.462
	Posttest	0.848	0.360	0.917	0.277
	Change	0.128	0.443	0.220	0.459
5. Where is the Sun located relative to the Milky Way?	Pretest	0.230	0.422	0.147	0.356
	Posttest	0.584	0.495	0.532	0.501
	Change	0.354	0.613	0.385	0.560
6. How are galaxies distributed throughout the universe?	Pretest	0.188	0.392	0.229	0.422
	Posttest	0.399	0.491	0.450	0.500
	Change	0.210	0.560	0.220	0.567
7. Which statement about observing stars in the Milky Way is most accurate?	Pretest	0.288	0.454	0.303	0.462
	Posttest	0.399	0.491	0.413	0.495
	Change	0.111	0.508	0.110	0.533
8. Which one statement about galaxy properties is correct?	Pretest	0.957	0.204	0.963	0.189
	Posttest	0.964	0.188	0.972	0.164
	Change	0.007	0.256	0.009	0.254
9. Which statement most accurately describes the contents of the Milky Way?	Pretest	0.820	0.385	0.917	0.277
	Posttest	0.877	0.330	0.927	0.262
	Change	0.057	0.378	0.009	0.347
10. Which one of the following is true about planets and stars in the Milky Way?	Pretest	0.317	0.467	0.385	0.489
	Posttest	0.471	0.501	0.532	0.501
	Change	0.154	0.466	0.147	0.448

Table 16.4. Mean fraction of correct pretest and posttest answers associated with the galaxies video lectures

Question	Spring 2012 vs. Spring 2013				
	***df*2**	***F***	***p***	**η^2**	**ω^2**
1. How many galaxies exist in the universe today?	245	0.644	0.423	0.003	−0.001
2. Relative to the Milky Way galaxy, the solar system is located:	245	0.453	0.502	0.002	−0.002
3. How many distinct shapes do galaxies have?	245	6.220	0.013*	0.025	0.021
4. Where in the universe is the Milky Way located today?	245	2.829	0.094	0.011	0.007
5. Where is the Sun located relative to the Milky Way?	244	0.213	0.645	0.001	−0.003
6. How are galaxies distributed throughout the universe?	245	0.019	0.889	0.000	−0.004
7. Which statement about observing stars in the Milky Way is most accurate?	245	0.000	0.983	0.000	−0.004
8. Which one statement about galaxy properties is correct?	245	0.003	0.953	0.000	−0.004
9. Which statement most accurately describes the contents of the Milky Way?	245	1.092	0.297	0.004	0.000
10. Which one of the following is true about planets and stars in the Milky Way?	245	0.008	0.927	0.000	−0.004

Table 16.5. Tests of significant gains associated with the galaxies video lectures

16.3 Test of the video lecture data

16.3.1 Data replacement for one black holes video posttest question

Between the Spring 2012 and the Spring 2013 semester tests, student performance on the black holes video lecture tests was not significantly different. Given that the pretest scores on Question 5, between the two semesters, were essentially the same (0.076 vs. 0.075), one may reasonably assume that test performance on Question 5 for students in the Spring 2013 semester would have been essentially the same as test performance on the same question for students in the Spring 2012 semester. One option is to use the Spring 2013 pretest and posttest data, as is. Another option is to replace the Question 5 posttest data on black holes for the Spring 2013 with the *mean* of the Question 5 posttest data on black holes for the Spring 2012 semester. The replacement serves the purpose of estimating the true data that I would obtain if we had not run out of time at the end of the black holes video lecture in the Spring 2013 semester. Hence, the aforementioned replacement, along with all other test data from the Spring 2012 and Spring 2013 semesters, constitutes my modified data set and thus provides a more accurate overall representation on student performance on the tests in Spring 2013.

16.3.2 Descriptive test statistics

Table 16.6 presents the overall semester performance and standard deviation on the black holes and galaxies pretest and posttest for the Spring 2012 and Spring 2013 semesters, using the black holes Question 5 posttest score replacement just discussed in Section 16.3.1. Table 16.6 also lists the number of students who submitted valid responses to the pretest and posttest questions associated with a particular video.

		Black Holes			
		Pretest		Posttest	
Semester	***n***	**Mean**	**Std. Dev**	**Mean**	**Std. Dev.**
Spring 2012	144	0.511	0.147	0.664	0.164
Spring 2013	133	0.489	0.163	0.645	0.132

		Galaxies			
		Pretest		Posttest	
Semester	***n***	**Mean**	**Std. Dev**	**Mean**	**Std. Dev.**
Spring 2012	138	0.537	0.170	0.674	0.199
Spring 2013	109	0.542	0.138	0.683	0.177

Table 16.6. Black holes and galaxies mean test gains

16.3.3 MANOVA test results

The data in Table 16.6 suggests that there does not seem to be any significant difference in overall performance when the order of video lecture instruction is re-arranged. To quantify the significance, I performed a Multivariate ANalysis Of VAriance (MANOVA) test on the combined Spring 2012 and Spring 2013 pretest and posttest data for both galaxies and black holes, of which instruction in only the Spring 2013 galaxies video lecture were presented in the conventional order. Recall that the subtopic associated with Question 5, that black holes do not suck, which appears at the end of the alternate video lecture, was not presented to the students in the Spring 2013 semester. Hence, I am using the modified data set, as outlined in Section 16.3.1.

Table 16.7 summarizes the significance of the overall effects between various aspects of the data. The *Topic* label refers to whether or not there is a difference in student performance on the black holes video lecture vs. the galaxies video lecture. The *Test* label refers to whether or not there is a significant difference between a pretest and the corresponding posttest. The *Semester* label refers to whether or not the data are significantly different between the semesters. The final sample sizes for the Spring 2012 and Spring 2013 semesters are, respectively, 114 and 94. Wilks' λ, a common measure in multivariate analysis (Everitt & Dunn, 1991), is used to measure the proportion of variance in the dependent variables that does not arise from each of the effect(s) under consideration. Values of λ range from 0 to 1, where $\lambda = 1$ represents groups with the same means, and $\lambda \to 0$ represents groups with different means.

Effect	**Wilks' λ**	**$F(1, 206)$**	**p**
Topic	0.960	8.64	0.004*
Topic * Semester	0.991	1.93	0.17
Test	0.415	290.5	<0.0005**
Test * Semester	0.999	0.297	0.59
Topic * Test	0.999	0.168	0.68
Topic * Test * Semester	1.000	0.001	0.97

Table 16.7. MANOVA test results for performance on the black holes and galaxies video lecture tests

The results of each effect, presented in Table 16.7, are discussed below:

1. The *Topic* effect is significant, which says that student performance on black holes is significantly different than on galaxies. As indicated in Table 16.6, student pretest and posttest scores on the galaxies topic are higher than those on the black holes topic.

2. The *Topic * Semester* effect is not significant, which says that the differences in scores between topics are not significant when further comparing the two semesters.

3. The *Test* effect is significant, which says that the differences between pretest and posttest scores are considered significant in general. This is an expected effect because students are expected to perform higher after instruction.

4. The *Test * Semester* effect is not significant, which says that the differences in pretest and posttest scores are not significantly different between semesters. This is, in part, a re-assuring check that because students in both semesters received the same black holes video lecture, their overall performance is not expected to

be significantly different. But because the *Test * Semester* effect is not significant, overall student performance on the galaxy tests is, likewise, not significantly different between the semesters. Therefore, the re-arrangement of instruction in the order suggested by my item response theory results is unlikely to produce any significant *overall* change in student performance on the topic. My results in Section 16.2.2, however, do support that the re-arrangement of topics may significantly influence performance on individual questions, but the influences still do not produce an overall significant change in test performance.

5. The *Topic * Test* effect is not significant, which says that the scores on either the black holes pretest or the black holes posttest are not significantly different than those on the equivalent galaxies test.

6. The *Topic * Test * Semester* effect is not significant, which says that for each topic, the pretest and posttest scores for either semester are not significantly different than the same scores in the other semester. Conceptually, that the *Topic * Test * Semester* effect is not significant means that there is no particular test whose scores stand out significantly more than scores on any of the other tests.

16.4 Discussion of results

The results presented in this Chapter can be summarized as follows:

1. The act of assigning students essentially at random to alternate classrooms to learn about black holes has no significant effect on the performance that the students would have if they learned the same material in their customary lecture classroom.

2. Pre-recorded instruction on the hardest black hole concept, to address the misconception that black holes do not suck, is significantly better than no instruction on it at all.

3. The re-arrangement of galaxy concepts, in the order suggested by my item response theory results, within pre-recorded presentations produces no statistically significant improvement in the gain on the tests associated with the galaxies presentations.
4. The re-arrangement of topics in the order suggested by my item response theory results, however, does significantly influence student performance on *individual test questions* affected by the topic rearrangement.

CHAPTER 17

CONCLUSIONS

17.1 Summary

In this dissertation, I have presented a comprehensive analysis of a new survey-like instrument (the ABI) consisting of 215 statements, in which each statement targets a specific common misconception in astronomy. I have shown that the instrument may be used as a retrospective survey of one's own past misconceptions, with the reliability of the data on par with that of standardized examinations. In Chapter 2, I outlined the teaching pedagogies of two veteran astronomy instructors at the University of Maine, and I have used the ABI to collect data from students taught under both instructors, each utilizing a different teaching pedagogy to teach essentially the same material. Brief results from each Chapter are summarized below:

1. In Chapter 3, I demonstrated that there are no statistically significant issues with statement misinterpretation or the correctness of response codes to the associated statements. I also examined the effect of fatigue in the process of completing the ABI and I found that student fatigue, if present, has no significant influence on students' self-reports. Hence, interested researchers need not concern themselves with the high number of ABI items (215). I also showed that addressing misconceptions is more effective in enabling students to reduce the number of misconceptions they endorse.

2. In Chapters 3-4, I demonstrated that the ABI is a reasonably valid instrument at assessing student misconceptions when compared against other instruments, such as multiple-choice exams.

3. In Chapters 5 and 7, I stated that the overall reported degree of misconception retainment, from the ABI, suffers significant selection effects; however, I

demonstrated that these effects have no significant influence on *correlations* between misconceptions.

4. In Chapters 8-12, I used principal components analysis to establish groups of the most correlated misconceptions. I found that some misconceptions are best unlearned by teaching in the context of misconceptions about multiple objects together (e.g., the Earth and the Moon), while others are best unlearned by not combining misconceptions about different objects (e.g., the Sun and other stars). I have also introduced a method to reduce the original set of 215 statements to four groups consisting in total of 27 statements. The four groups project the same misconception themes as those resulting from performing principal components analysis on predetermined subsets of the data.
5. In Chapters 13-15, I used the methodology of item response theory to present the most logical order to teach all statements within a given topic. I presented unique orders to sequence instruction as a function of the age of the audience.
6. In Chapter 16, I designed and analyzed the results of video lecture tests in a kind of partial crossover design, with topics in the video lectures specifically arranged to test the predictions of the results in Chapter 14. In general, I found that the re-arrangement of topics had no significant effect on overall student test gain. On the other hand, I found that the re-arrangement of topics in the order suggested by my item response theory results may have influenced student performance on *individual test questions*.

17.2 Contribution

The astronomy education community is well aware that misconceptions constitute deep-seated beliefs that often interfere with learning, and that misconceptions in astronomy are prevalent even among college students. The most important contribution of my overall research project to the field of astronomy education is that of a retrospective survey to assess, as directly as possible, misconceptions held by these students. My analysis in Chapter 4 shows that such a survey *can* be used to gather meaningful data about the students' own misconceptions. It is my anticipation that the astronomy education community will find such an instrument, or a variation of it, fruitful in their research.

17.3 Future work

From over 20 years of his own teaching experience, NFC has identified 1700 misconceptions in astronomy (Comins, 2001, 2014). That so many of these misconceptions have not only been identified but continue to be endorsed today by college students strongly suggests that there is no shortage of opportunity for members of the astronomy education community to initiate a research project stemming from the misconceptions survey.

Data from my overall research project was conducted exclusively within the University of Maine. One suggestion for future work would be to administer the ABI to college students taking introductory astronomy at other universities. A nationwide implementation of the ABI would allow for substantially faster data acquisition. The results of performing principal components analysis on data from a nationwide administration of the ABI could conveniently fine-tune the results of this research project with regard to the misconceptions that are the most highly correlated with each other.

A second suggestion for future work would be to administer a subset of ABI items (e.g., those under "Galaxies"), to children and adolescents directly. While college students seem to handle a survey of 215 statements, such a survey is understandably daunting for children, so one could choose to analyze just a small part of the inventory.

A further suggestion would be to administer the ABI (or a subset of it) at the beginning of the semester, as a pretest and before any instruction, and then re-administer the ABI as usual at the end of the course. The advantage of such a design would be to make comparisons in the style of pretest vs. posttest, using the *same* instrument, rather than using other instruments to supplement for the pretest. One way to control for pretest-posttest effects would be to have half of the class take the pretest and the other half take the posttest, with the students randomly selected for each test.

One final suggestion for future work would be to consider the effect of ABI scores on substantially different teaching pedagogies. In their courses, NFC lectures with targeted emphasis on misconceptions, while Prof. David J. Batuski lectures with emphasis on clicker questions and homework questions. Neither instructor teaches with emphasis on in-class tutorials or questions in the style of “think-pair-share” (Slater & Adams, 2002). It would be interesting to see the effects of other teaching pedagogies on the long-term disambiguation of misconceptions.

REFERENCES

Anderson, C. W., & Smith, E. L. (1988). *Children's conceptions of light and color: Understanding the role of unseen rays* (Tech. Rep. No. 166). East Lansing, MI: Michigan State University, College of Education, Institute for Research on Teaching, Res. Series.

Arny, T. T. (2004). *Explorations: Stars, galaxies, and planets*. Boston: McGraw-Hill.

Arny, T. T. (2006). *Explorations: an introduction to astronomy* (4th ed.). Boston: McGraw-Hill.

Astronomy Education Review. (2013). *Browse - Astronomy Education Review.* Retrieved August 21, 2013, from http://aer.aas.org/resource/1/aerscz/.

Bailey, J. M., Prather, E. E., Johnson, B., & Slater, T. F. (2009). College students' preinstructional ideas about stars and star formation. *Astronomy Education Review*, *8*, 010110.

Bailey, J. M., & Slater, T. F. (2004). A review of astronomy education research. *Astronomy Education Review*, *2*(2), 20-45.

Baker, F. (2001). *The basics of item response theory*. College Park, MD: University of Maryland.

Beatty, J. K., Petersen, C. C., & Chaikin, A. (1999). *The new solar system* (4th ed.). Cambridge: Cambridge University Press.

Becker, L. A. (1999). *Measures of effect size (strength of association).* Retrieved May 10, 2014, http://www.uccs.edu/lbecker/glm_effectsize.html.

Bennett, J., Donahue, M., Schneider, N., & Volt, M. (2012). *The cosmic perspective* (6th ed.). Boston: Addison–Wesley.

Borkowski, J. (2014). *Tests for homogeneity of variance.* Retrieved January 25, 2014, http://www.math.montana.edu/~jobo/st541/sec2e.pdf.

Bransford, J. D. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brewin, C. R., Andrews, B., & Gotlib, I. H. (1993). Psychopathology and early experience: A reappraisal of retrospective reports. *Psychological Bulletin*, *113*(1), 82-98.

Cai, L. (2008). SEM of another flavor: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, *61*, 309-329.

Carroll, S. M. (2003). *Spacetime and geometry: An introduction to general relativity*. Boston: Addison-Wesley.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Center for Astronomy Education. (2013). *astro101: Publications.* Retrieved August 21, 2013, from http://astronomy101.jpl.nasa.gov/publications/.

Chaisson, E., & McMillan, S. (1996). *Astronomy today* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Chajewski, M., & Lewis, C. (2009). Optimizing item exposure control algorithms for polytomous computerized adaptive tests with restricted item banks. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.*

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning. *American Educator*, 6-11.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*, 66-71.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385-396.

Comins, N. F. (2001). *Heavenly errors: Misconceptions about the real nature of the universe*. New York: Columbia University Press.

Comins, N. F. (2008). *Astronotes*. New York: Freeman Custom Publishing.

Comins, N. F. (2013). *Discovering the essential universe* (5th ed.). New York: W. H. Freeman & Company.

Comins, N. F. (2014). *Heavenly errors.* Retrieved January 24, 2014, http://www.physics.umaine.edu/ncomins/.

Comins, N. F., & Kaufmann III, W. J. (2012). *Discovering the universe* (9th ed.). New York: W. H. Freeman.

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885-901.

diSessa, A. A. (1982). Unlearning aristotelian physics: A study of knowledge based learning. *Cognitive Science*, *6*, 37-75.

Dixon, R. T. (1992). *Dynamic astronomy*. Englewood Cliffs, NJ: Prentice Hall.

Driver, R. (1983). *The pupil as scientist?* Milton Keynes, England: Open University Press.

DuToit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.

Eaton, J. F., Anderson, C., & Smith, E. (1983). When students don't *know* they don't know. *Science and Children*, *20*(7), 6-9.

Edelen, M. O., & Reeve, B. R. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual. Life Res.*, *16*(1), 5-18.

Einasto, M., Liivamägi, L. J., Saar, E., Einasto, J., Tempel, E., Tago, E., & Martínez, V. J. (2011). SDSS DR7 superclusters: Principal component analysis. *Astronomy & Astrophysics*, *535*, A36.

Eldep Jr., G. H., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival data: Studying lives*. Newbury Park, CA: Sage Publication.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Engelbrektson, S. (1994). *Astronomy: Through space and time*. Dubuque, IA: Wm. C. Brown.

Everitt, B. S., & Dunn, G. (1991). *Applied multivariate data analysis* (1st ed.). London: Edward Arnold.

Eylon, B., & Linn, M. C. (1988). Learning and instruction: An examination of four research perspectives in science education. *Review of Educational Research*, *58*(3), 251-301.

Favia, A., Comins, N. F., Thorpe, G. L., & Batuski, D. J. (2014). A direct examination of college student misconceptions in astronomy: A new instrument. *Journal and Review of Astronomy Education and Outreach*, *1*, A21-A39.

Fix, J. (2004). *Astronomy: Journey to the cosmic frontier* (3rd ed.). New York: McGraw–Hill.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development*, *51*(1), 1-87.

Freedman, R., Geller, R., & Kaufmann III, W. J. (2011). *Universe* (9th ed.). New York: W. H. Freeman & Company.

Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology*, *21*(4), 572-583.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377-393.

Gray, P. A. (2006). *Gender differences in science misconceptions in eighth grade astronomy*. Unpublished doctoral dissertation, Widener University.

Hartmann, W. (1999). *Moons & planets* (4th ed.). Beverly, MA: Wadsworth.

Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the 'remembrance of things past': A longitudinal evaluation of the retrospective method. *Psychological Assessment*, *6*, 92-101.

Hester, J., Blumenthal, G., Smith, B., Burstein, D., Greeley, R., & Voss, H. (2007). *21st century astronomy* (2nd ed.). New York: W. W. Norton & Company.

Hole, G. (2013). *Testing for homogeneity of variance with Hartley's* F_{max} *test.* Retrieved January 25, 2014, http://www.sussex.ac.uk/Users/grahamh/RM1web/Testing%20for%20homogeneity%20of%20variance.pdf.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Humphreys, L. G., & Montanelli Jr., R. G. (1975). Multivariate behavioral research: An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*(2), 193-205.

Kabacoff, R. I. (2012). *Quick-R: Factor Analysis.* Retrieved July 18, 2013, http://www.statmethods.net/advstats/factor.html.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kaler, J. B. (1994). *Astronomy!* New York: Harper Collins.

Kaufmann III, W. J., & Freedman, R. (2002). *Universe* (6th ed.). New York: W. H. Freeman & Company.

Kempton, W. (1986). Two theories of home heat control. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (p. 222-242). Cambridge: Cambridge University Press.

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation.* Thousand Oaks, CA: Sage Publications.

Koupelis, T. (2014). *In quest of the universe* (7th ed.). Sudbury, MA: Jones & Bartlett Learning, LLC.

Koupelis, T., & Kuhn, K. F. (1999). *In quest of the universe* (5th ed.). Sudbury, MA: Jones & Bartlett Learning, LLC.

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *Journal of Neuroscience*, *25*(11), 2977-2982.

Lacey, M. (1998). *Statistical topics.* Retrieved March 24, 2014, http://www.stat.yale.edu/Courses/1997-98/101/stat101.htm.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, *12*(2), 1-11.

Lee, K., & Ashton, M. C. (2007). Factor analysis in personality research. In R. W. Robins (Ed.), *Handbook of research methods in personality psychology* (p. 424-443). New York: Guilford Press.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*. New York: Springer-Verlag.

Lewis, J. (2004). *Physics and chemistry of the solar system* (2nd ed.). Salt Lake City: Academic Press.

Libarkin, J. C., Asghar, A., Crockett, C., & Sadler, P. (2011). Invisible misconceptions: Student understanding of ultraviolet and infrared radiation. *Astronomy Education Review*, *10*, 010105.

Lombardi, D., Sinatrab, G. M., & Nussbaum, E. M. (2013). Plausibility reappraisals and shifts in middle school students' climate change conceptions. *Learning and Instruction*, *27*, 50-62.

LoPresto, M. C., & Murrell, S. R. (2011). An astronomical misconceptions survey. *Journal of College Science Teaching*, *40*(5), 14-23.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Earlbaum.

Miller, L. A., Lovler, R. L., & McIntire, S. A. (1999). *Foundations of psychological testing: A practical approach* (4th ed.). Thousand Oaks, CA: SAGE.

Moreno, R. (2004). Decreasing cognitive load in novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*(1), 99-113.

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (p. 407-423). New York: Guilford.

Neuhaus, J. O., & Wrigley, C. (1954). The quartimax method: An analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, *7*, 81-91.

Olson, J. M., & Cal, A. V. (1984). Source credibility, attitudes, and the recall of past behaviors. *European Journal of Social Psychology*, *14*, 203-210.

Palen, S., Lay, L., Smith, B., & Blumenthal, G. (2011). *Understanding our universe*. New York: W. W. Norton & Company.

Pasachoff, J. M. (1998). *Astronomy: From the earth to the universe* (5th ed.). Orlando, FL: Saunders/Harcourt College Publishing.

Pasachoff, J. M., & Filippenko, A. (2007). *The cosmos: Astronomy in the new millennium* (3rd ed.). Stamford, CT: Cengage Learning.

Posner, G. J., Strike, K. A., & Hewson, P. W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211-227.

Prather, E. E., & Brissenden, G. (2009). Clickers as data gathering tools and students: Attitudes, motivations, and beliefs on their use in this application. *Astronomy Education Review*, *8*, 010103.

Prather, E. E., Slater, T. F., Adams, J. P., Bailey, J. M., Jones, L. V., & Dostal, J. A. (2004). Research on a lecture-tutorial approach to teaching introductory astronomy for non-science majors. *Astronomy Education Review*, *3*, 122-136.

Prather, J. P. (1985). Philosophical examination of the problem of the unlearning of incorrect science concepts. *ERIC*.

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (p. 55-73). New York: Oxford University Press.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292.

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, *35*(3), 265-296.

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The Astronomy and Space Science Concept Inventory: Development and validation of assessment instruments aligned with the K-12 National Science Standards. *Astronomy Education Review*, *8*, 010111.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Tech. Rep. No. 17). Fredericton: University of New Brunswick.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350-353.

Schneider, S., & Arny, T. (2007). *Pathways to astronomy*. New York: McGraw–Hill.

Seeds, M. (2010). *Horizons: Exploring the universe* (10th ed.). Belmont, CA: Brooks Cole.

Shipman, H. (1978). *Restless universe: Introduction to astronomy*. Boston: Houghton Mifflin Harcourt.

Shu, F. (1982). *The physical universe: An introduction to astronomy*. Mill Valley, CA: University Science Books.

Slater, T. F. (2013). *Educational theory underlying astronomy clicker questions & peer instruction.* Retrieved May 17, 2014, http://astronomyfacultylounge.wordpress.com/2013/01/02/educational-theory-underlying-astronomy-clicker-questions-peer-instruction/.

Slater, T. F., & Adams, J. P. (2002). *Learner-centered astronomy teaching: Strategies for astro 101*. Boston: Addison-Wesley.

Slater, T. F., & Freedman, R. (2012). *Investigating astronomy: A conceptual view of the universe* (2nd ed.). New York: W. H. Freeman & Company.

Small, T. A., Ma, C., Sargent, W. L. W., & Hamilton, D. (1998). The Norris Survey of the Corona Borealis supercluster. III. Structure and mass of the supercluster. *The Astrophysical Journal*, *492*, 45-56.

Smith III, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115-163.

Snow, T. (1990). *Essentials of the dynamic universe* (4th ed.). Eagan, MN: West Publishing Company.

Snow, T., & Brownsberger, K. (1997). *Universe: origins and evolution*. Los Angeles: West Group.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive science, and educational theory and practice* (p. 147-176). Albany, NY: State University of New York Press.

Svedham, H. (2011). Introduction to the special issue on comparative planetology. *Planetary and Space Science*, *59*, 887-888.

Taylor, F. W. (2011). Comparative planetology, climatology and biology of venus, earth and mars. *Planetary and Space Science*, *59*(10), 889-899.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, *1*, 35-39.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*, 45-69.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535-585.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (p. 3-34). New York: Routledge.

Walker, J. T. (1985). *Using statistics for psychological research: An introduction.* New York: CBS College Publishing.

Wallace, C. S., & Bailey, J. M. (2010). Do concept inventories actually measure anything? *Astronomy Education Review*, *9*, 010116.

Wallace, C. S., Prather, E. E., & Duncan, D. K. (2011). A study of general education astronomy students' understandings of cosmology. Part I. Development and validation of four conceptual cosmology surveys. *Astronomy Education Review*, *10*, 010106.

White, B. Y. (1982). Sources of difficulty in understanding newtonian dynamics. *Cognitive Science*, *7*, 41-65.

Zeilik, M. (1997). *Astronomy: The evolving universe* (8th ed.). Hoboken, NJ: Wiley.

Zeilik, M. (2002). *Astronomy: The evolving universe* (9th ed.). Cambridge University Press.

APPENDIX A

THE 215 STATEMENTS OF THE STUDY AND THEIR MEAN MISCONCEPTION RETAINMENT SCORE FROM FALL 2009 TO FALL 2013

Stars:

1.	sA1.	all of the stars were created at the same time	1.60
2.	sA2.	there are 12 zodiac constellations	2.13
3.	sA3.	all of the stars are about as far away from the Earth as the Moon	1.62
4.	sA4.	all stars are white	1.52
5.	sA5.	the constellations are only the stars we connect to make patterns	2.28
6.	sA6.	we are looking at stars as they are now	1.66
7.	sA7.	stars actually twinkle — change brightness	2.02
8.	sA8.	the north star is the brightest star in the sky	2.03
9.	sA9.	stars have spokes	1.90
10.	sA10.	all stars have planets	1.80
11.	sA11.	stars last forever	1.48
12.	sA12.	the brighter a star is, the hotter it is	2.33
13.	sA13.	all stars are evenly distributed on the celestial sphere	1.89
14.	sA14.	all stars are the same distance from the Earth	1.45
15.	sA15.	all stars have same color and size	1.48
16.	sA16.	pulsars are pulsating stars	2.36
17.	sA17.	all stars are smaller than the Sun	1.62
18.	sA18.	the galaxy, solar system and universe are the same things	1.46

19.	sA20.	stars just existed — they don't make energy or change size or color	1.65
20.	sA21.	all stars end up as white dwarves	2.04
21.	sA22.	all stars are stationary — fixed on the celestial sphere	1.92
22.	sA23.	stars emit only one color of light	1.79
23.	sA24.	stars are closer to us than the Sun	1.69
24.	sA25.	there are exactly 12 constellations	1.70
25.	sA27.	all the stars in an asterism move together	2.40
26.	sA28.	a nova is the most powerful explosion	2.04
27.	sA29.	stars in the Milky Way are as close to each other as planets are to the Sun	1.89
28.	sA30.	stars run on fuel: gasoline or natural gas	1.84
29.	sA31.	"metals" have always existed in the universe	2.29
30.	sA32.	stars follow you in your car	1.44
31.	sA33.	we see the same constellations at night throughout the year	1.67
32.	sA34.	stars are fixed in space	1.72
33.	sA35.	stars in a binary system (two stars bound together by their gravity) would quickly collide	2.15
34.	sA37.	all stars are isolated from all other stars (none are binary)	1.92

Solar System, Misc.:

35.	sA40.	the asteroid belt is an area like we see in star wars, very densely packed	2.08
36.	sA41.	Mercury is so named because there is much mercury on it	1.71
37.	sA42.	comet tails are burning — because the comet is moving so fast	1.99
38.	sA43.	there is plant life on other planets in our solar system	1.72

39.	sA44.	Pluto is always farther from the Sun than is Neptune	2.10
40.	sA45.	a shooting star is actually a star whizzing across the universe or falling through the sky	1.80
41.	sA46.	Jovian planets (Jupiter, Saturn, Uranus, Neptune) have solid surfaces	1.85
42.	sA47.	the asteroid belt is between Earth and Mars	2.02
43.	sA48.	the Solar System is the whole universe or the whole galaxy	1.59
44.	sA49.	Jupiter is almost large and massive enough to be a star	2.18
45.	sA50.	all orbits around Sun are circular	1.71
46.	sA51.	planets revolve around the Earth	1.50
47.	sA52.	all planets orbit exactly in the plane of the ecliptic	2.02
48.	sA53.	Pluto is a large, jovian (Jupiter-like) planet	1.60
49.	sA54.	all constellations look like things they are named for	1.88
50.	sA56.	comets last forever	1.72
51.	sA57.	each planet has one moon	1.53
52.	sA58.	Mercury (closest planet to the Sun) is hot everywhere on its surface	2.03
53.	sA59.	the day on each planet is 24 hours long	1.54
54.	sA60.	all stars have prograde rotation (spin same way as the Earth)	1.74
55.	sA62.	there are no differences between meteors, meteorites, meteoroids	1.74
56.	sA63.	asteroids, meteoroids, comets are same	1.65
57.	sA66.	optical telescopes are the only “eyes” astronomers have on the universe	1.84
58.	sA67.	humans have never landed a spacecraft on another planet	1.70
59.	sA68.	we do not have telescopes in space	1.59
60.	sA69.	all planets have been known for hundreds of years	1.88

61.	sA70.	comets are molten rock hurtling through space at high speeds and their tails are jet wash behind them	2.05
62.	sA72.	there are many galaxies in a solar system	1.92
63.	sA75.	comets are solid, rocky debris	2.13
64.	sA76.	Jupiter’s great red spot is a volcano erupting	1.85
		Moon:	
65.	sA77.	there is only one moon — ours	1.34
66.	sA78.	the Moon doesn’t cause part of the tides	1.55
67.	sA79.	we see all sides of the Moon each month	1.78
68.	sA80.	craters are volcanic in origin	1.92
69.	sA83.	the Moon is at a fixed distance from Earth	1.96
70.	sA84.	the Moon changes physical shape throughout its cycle of phases	1.63
71.	sA85.	the Moon doesn’t rotate since we see only one side of it	1.83
72.	sA87.	the Moon has seas and oceans of water	1.64
73.	sA88.	the Moon is older than the Earth: a dead planet that used to be like Earth	1.80
74.	sA89.	the Moon is about the same temperature as the Earth	1.61
75.	sA90.	the Moon has a helium atmosphere	1.97
76.	sA91.	the Moon has an atmosphere like the Earth	1.65
77.	sA92.	the Moon has a smooth surface	1.57
78.	sA93.	the Moon sets during daylight hours and is not visible then	1.61
79.	sA94.	there is a real man in the Moon	1.38
80.	sA96.	because the Moon reflects sunlight, it has a mirror-like surface	2.00
81.	sA97.	the Moon will someday crash into Earth	1.91

82.	sA98.	the Moon is a captured asteroid	2.05
83.	sA99.	a lunar month is exactly 28 days long	2.47
84.	sA100.	at new Moon we are seeing the “far side” of the Moon	2.04
85.	sA102.	the Moon follows you in your car	1.42
86.	sA103.	the Moon is larger at the horizon than when it is overhead	2.23
87.	sA104.	the side of the moon we don’t see is forever “dark”	2.04
88.	sA105.	the moon is lit by reflected “Earth light” (that is, sunlight scattered off the Earth toward the Moon)	2.01

Venus:

89.	sA106.	life as we know it can exist on Venus	1.75
90.	sA107.	clouds on Venus are composed of water, like clouds on earth	1.93
91.	sA108.	Venus is very different from earth in size	1.97
92.	sA109.	Venus is a lot like the Earth in temperature	1.85
93.	sA110.	Venus is always the first star out at night	2.10

Earth:

94.	sA111.	Earth’s axis is not tilted compared to the ecliptic	1.86
95.	sA112.	summer is warmer because we are closer to the sun during the summertime	2.01
96.	sA113.	once ozone is gone from the Earth’s atmosphere, it will not be replaced	2.45
97.	sA114.	Earth and Venus have similar atmospheres	2.00
98.	sA115.	Earth is at the center of the universe	1.49
99.	sA116.	Earth is the biggest planet	1.44
100.	sA118.	Spring Tide is in the spring	2.31
101.	sA122.	X-rays can reach the ground	1.99

102.	sA125.	meteoroids enter the atmosphere a few times a night	2.20
103.	sA126.	you can see a solar eclipse from anywhere on Earth that happens to be facing the Sun at that time	2.20
104.	sA127.	auroras are caused by sunlight reflecting off polar caps	2.21
105.	sA128.	the Moon is not involved with any eclipses	1.58
106.	sA129.	the day has always been 24 hours long	2.14
107.	sA130.	the air is a blue gas	1.64
108.	sA131.	Halley's comet will eventually hit Earth	2.17
109.	sA133.	the sun orbits the Earth	1.45
110.	sA135.	solar eclipses happen about once a century and are seen everywhere on Earth	1.97
111.	sA137.	only Earth among the planets and moons has gravity	1.69
112.	sA141.	seasons were chosen haphazardly	2.12
113.	sA142.	meteorites have stopped falling onto the Earth	1.79
114.	sA143.	the Earth will last forever	1.48
115.	sA144.	the Earth's magnetic poles go through its rotation poles	2.30
116.	sA145.	planes can fly in space	1.66
117.	sA146.	a day is exactly 24 hours long	1.89
118.	sA147.	a year is exactly 365 days long	1.74
119.	sA148.	seasons are caused by speeding up and slowing down of Earth's rotation	1.81
120.	sA149.	the Earth orbits the sun at a constant speed	2.38
121.	sA150.	the Earth is in the middle of the Milky Way galaxy	1.72
122.	sA151.	the sky is blue because it reflects sunlight off oceans and lakes	1.89
123.	sA152.	the Earth is the only planet with an atmosphere	1.61
124.	sA153.	comets affect the weather	2.00

125.	sA154.	the Earth is not changing internally	1.94
126.	sA156.	the tides are caused just by the Earth’s rotation	1.70
127.	sA157.	Earth has a second moon that only comes around once in awhile — “once in a blue moon”	1.64
128.	sA158.	the Sun is directly overhead everywhere on Earth at noon	1.84
129.	sA159.	tides are caused just by ocean winds	1.57
130.	sA160.	the Earth is flat	1.50
		Mars:	
131.	sA161.	Mars is green (from plant life)	1.62
132.	sA164.	Mars has running water on its surface now	1.78
133.	sA165.	Mars could be made inhabitable	2.29
134.	sA166.	Mars is the second largest planet	1.71
135.	sA167.	life, when it did exist on Mars, was quite advanced	1.68
136.	sA168.	there are Lowellian canals on Mars built by intelligent beings	1.73
137.	sA169.	Mars is Hot because it is red ... Mars — god of fire	1.61
138.	sA170.	Mars is the sister planet to earth in physical properties and dimensions	2.22
		Saturn:	
139.	sA171.	Saturn is the only planet with rings	1.59
140.	sA172.	Saturn’s rings are solid	1.67
141.	sA174.	Saturn’s rings are caused by the planet spinning so fast	1.96
142.	sA176.	Saturn has only one ring	1.64

Sun:

143.	sA177.	the Sun is a specific type of astronomical body with its own properties. It is not a star	1.45
144.	sA178.	the Sun will burn forever	1.52
145.	sA180.	the Sun is the hottest thing in the galaxy	1.76
146.	sA181.	the Sun does not move through space	2.04
147.	sA182.	the Sun does not cause part of the tides	2.08
148.	sA183.	sunspots are hot spots on the Sun’s surface	2.24
149.	sA184.	the Sun will blow up, become a black hole, and swallow the earth	1.98
150.	sA185.	the Sunspot cycle is 11 years long	2.54
151.	sA186.	the Sun’s surface temperature is millions of degrees Fahrenheit	2.43
152.	sA187.	Sunspots are constant fixtures on the sun	1.97
153.	sA188.	the Sun is yellow	1.90
154.	sA189.	the Sun is the brightest star in universe	1.65
155.	sA190.	the Sun is the brightest object in the universe	1.77
156.	sA191.	the Sun always sets due west	2.44
157.	sA192.	the Sun is made of fire	1.47
158.	sA193.	the Sun is a “heat planet”	1.69
159.	sA196.	the Sun is the smallest star in universe	1.73
160.	sA197.	the Sun has no atmosphere	2.15
161.	sA198.	the Sun is the largest star	1.65
162.	sA199.	the Sun is hottest on its surface	2.02
163.	sA200.	the Sun has a solid core	2.16
164.	sA201.	the Sun has only a few percent of the mass in the solar system	2.21

165.	sA202.	the Sun is mostly iron	2.05
166.	sA204.	the Sun’s surface is perfectly uniform	1.79
167.	sA206.	the entire Sun is molten lava	1.59
168.	sA208.	the Sun will explode as a nova	2.38
169.	sA209.	the Sun is hottest star	1.68
170.	sA211.	it is possible that the Sun could explode in the “near future”	1.92
171.	sA213.	the Sun doesn’t rotate	1.93
172.	sA214.	the Sun is the only source of light in the galaxy — Sunlight reflects off planets and stars so we can see them.	1.77
173.	sA215.	Sunspots are where meteors crash into the Sun	1.89
174.	sA217.	it is more dangerous to look at the Sun during an eclipse because the radiation level from sun is greater then, than when there is no eclipse	2.22

Galaxies:

175.	sA218.	the Milky Way is the only galaxy	1.43
176.	sA219.	the solar system is not in the Milky Way (or any other) galaxy	1.66
177.	sA220.	all galaxies are spiral	1.87
178.	sA221.	the Milky Way is the center of the universe	1.76
179.	sA222.	the Sun is at the center of the Milky Way galaxy	1.89
180.	sA224.	the Sun is at the center of the universe	1.63
181.	sA225.	there are only a few galaxies	1.72
182.	sA226.	the galaxies are randomly distributed	2.46
183.	sA227.	we see all the stars that are in the Milky Way	1.86
184.	sA228.	all galaxies are the same in size and shape	1.75

185.	sA230.	the Milky Way is just stars — no gas and dust	1.73
186.	sA231.	new planets and stars don’t form today	1.81

Black Holes:

187.	sA232.	black holes create themselves from nothing	1.89
188.	sA233.	black holes last forever	2.22
189.	sA234.	black holes really don’t exist	1.76
190.	sA235.	black holes are empty space	2.01
191.	sA237.	black holes do not have mass	2.03
192.	sA238.	black holes are like huge vacuum cleaners, sucking things in	2.27
193.	sA240.	black holes are doors to other dimensions	1.79
194.	sA242.	black holes can be seen visually, like seeing a star or planet	2.05
195.	sA243.	we could live in a voyage through a black hole	1.71
196.	sA244.	we could travel through time in a black hole	1.82
197.	sA245.	black holes get bigger forever and nothing can stop them from doing so	2.08
198.	sA246.	black holes are actual holes in space	1.85
199.	sA247.	a single black hole will eventually suck in all the matter in the universe	1.89

General Astrophysics:

200.	sA248.	cosmic rays are light rays	2.28
201.	sA252.	astronomy and astrology are the same thing	1.62
202.	sA253.	gravity will eventually pull all the planets together	1.89
203.	sA254.	satellites need continuous rocket power to stay in orbit around the Earth	1.66
204.	sA255.	light travels infinitely fast	1.87

205.	sA256.	space is infinite	2.58
206.	sA258.	telescopes cannot see any details on any of the planets	1.80
207.	sA259.	gravity is the strongest force in the universe	2.33
208.	sA261.	we can hear sound in space	2.07
209.	sA262.	the universe as a whole is static (unchanging)	1.72
210.	sA263.	astronomical ideas of mass, distance, and temperature of planets are all speculative	2.37
211.	sA267.	there is a center to the universe	2.23
212.	sA270.	smaller telescopes enable astronomers to see smaller details	1.86
213.	sA271.	the most important function of a telescope is magnification	2.14
214.	sA272.	all space debris existing today is the result of planet collisions and explosions on planets	2.24
215.	sA273.	astronomers mostly work with telescopes	2.14

APPENDIX B

ONE RANDOMIZED INVENTORY OF THE 215 BELIEFS

1. sA157. Earth has a second moon that only comes around once in awhile — “once in a blue moon”
2. sA235. black holes are empty space
3. sA221. the Milky Way is the center of the universe
4. sA43. there is plant life on other planets in our solar system
5. sA80. craters are volcanic in origin
6. sA109n. Venus is hotter than the Earth
7. sA196. the Sun is the smallest star in universe
8. sA27. all the stars in an asterism move together
9. sA57. each planet has one moon
10. sA94. there is a real man in the Moon
11. sA181n. the Sun moves through space
12. sA17n. many stars are smaller than the Sun
13. sA165. Mars could be made inhabitable
14. sA176. Saturn has only one ring
15. sA4n. star have colors other than white
16. sA131n. Halley’s comet will never collide with Earth
17. sA49n. Jupiter is nowhere large or massive enough to be a star
18. sA258n. telescopes can see details on many of the planets
19. sA147. a year is exactly 365 days long
20. sA209n. other stars are hotter than the Sun
21. sA204. the Sun’s surface is perfectly uniform
22. sA32. stars follow you in your car
23. sA187. Sunspots are constant fixtures on the sun

24. sA228n. galaxies have a variety of sizes and shapes
25. sA10. all stars have planets
26. sA67n. humans have landed spacecraft on other planets
27. sA24n. the Sun is closer to us than the stars
28. sA102. the Moon follows you in your car
29. sA242n. black holes are invisible
30. sA70. comets are molten rock hurtling through space at high speeds and their tails are jet wash behind them
31. sA87. the Moon has seas and oceans of water
32. sA122n. X-rays are absorbed before they reach the ground
33. sA144. the Earth's magnetic poles go through its rotation poles
34. sA105. the moon is lit by reflected "Earth light" (that is, sunlight scattered off the Earth toward the Moon)
35. sA110. Venus is always the first star out at night
36. sA240n. black holes are not really doors to other dimensions
37. sA99. a lunar month is exactly 28 days long
38. sA76. Jupiter's great red spot is a volcano erupting
39. sA151. the sky is blue because it reflects sunlight off oceans and lakes
40. sA200n. the Sun's core is plasma
41. sA85n. the Moon rotates even though we see only one side of it
42. sA8. the north star is the brightest star in the sky
43. sA35n. stars in a binary system (two stars bound together by their gravity) do not collide
44. sA226. the galaxies are randomly distributed
45. sA47. the asteroid belt is between Earth and Mars
46. sA185. the Sunspot cycle is 11 years long
47. sA174. Saturn's rings are caused by the planet spinning so fast

48. sA118. Spring Tide is in the spring
49. sA21. all stars end up as white dwarves
50. sA273. astronomers mostly work with telescopes
51. sA30n. stars run on nuclear reactions instead of gasoline or natural gas
52. sA217. it is more dangerous to look at the Sun during an eclipse because the radiation level from sun is greater then, than when there is no eclipse
53. sA141n. seasons were chosen systematically
54. sA255n. light travels at a finite speed
55. sA62n. meteors, meteorites, and meteoroids mean different things
56. sA40. the asteroid belt is an area like we see in star wars, very densely packed
57. sA161. Mars is green (from plant life)
58. sA53. Pluto is a large, jovian (Jupiter-like) planet
59. sA91. the Moon has an atmosphere like the Earth
60. sA246. black holes are actual holes in space
61. sA263n. astronomical ideas of mass, distance, and temperature of planets are definite
62. sA111n. Earth's axis is tilted compared to the ecliptic
63. sA153. comets affect the weather
64. sA128n. the Moon is involved with eclipses
65. sA191. the Sun always sets due west
66. sA171. Saturn is the only planet with rings
67. sA218n. the Milky Way is one of many galaxies
68. sA177n. the Sun is a star
69. sA1n. all of the stars were created at different times
70. sA248. cosmic rays are light rays

71.	sA14n.	stars are at different distances from the Earth
72.	sA232.	black holes create themselves from nothing
73.	sA77.	there is only one moon — ours
74.	sA106n.	life as we know it would suffocate on Venus
75.	sA29.	stars in the Milky Way are as close to each other as planets are to the Sun
76.	sA254n.	satellites can stay in orbit around the Earth without continuous rocket power
77.	sA98.	the Moon is a captured asteroid
78.	sA137n.	Earth is one of many worlds among the planets and moons that has gravity
79.	sA60.	all stars have prograde rotation (spin same way as the Earth)
80.	sA225n.	there are billions of galaxies
81.	sA154n.	the Earth is changing internally
82.	sA168.	there are Lowellian canals on Mars built by intelligent beings
83.	sA199n.	the Sun is coldest on its surface
84.	sA272.	all space debris existing today is the result of planet collisions and explosions on planets
85.	sA184.	the Sun will blow up, become a black hole, and swallow the earth
86.	sA20n.	stars convert energy and change size or color
87.	sA7.	stars actually twinkle — change brightness
88.	sA46n.	Jovian planets (Jupiter, Saturn, Uranus, Neptune) have gaseous or icy surfaces
89.	sA84n.	the Moon's physical shape remains constant throughout its cycle of phases
90.	sA244.	we could travel through time in a black hole
91.	sA125.	meteoroids enter the atmosphere a few times a night

92. sA50n. few objects orbit the Sun in circular orbits
93. sA189. the Sun is the brightest star in universe
94. sA156. the tides are caused just by the Earth's rotation
95. sA88. the Moon is older than the Earth: a dead planet that used to be like Earth
96. sA261. we can hear sound in space
97. sA107. clouds on Venus are composed of water, like clouds on earth
98. sA11n. stars eventually die
99. sA12. the brighter a star is, the hotter it is
100. sA126. you can see a solar eclipse from anywhere on Earth that happens to be facing the Sun at that time
101. sA259. gravity is the strongest force in the universe
102. sA33n. we see different constellations at night throughout the year
103. sA231n. new planets and stars are forming today
104. sA188. the Sun is yellow
105. sA243n. a voyage through a black hole would be fatal
106. sA51. planets revolve around the Earth
107. sA230n. the Milky Way contains gas and dust
108. sA89n. the Moon and Earth have different temperatures
109. sA143. the Earth will last forever
110. sA48n. the Solar System is only part of our galaxy
111. sA256. space is infinite
112. sA202n. the Sun has very little iron in it
113. sA31. "metals" have always existed in the universe
114. sA100. at new Moon we are seeing the "far side" of the Moon
115. sA201n. the Sun has nearly all the mass in the solar system
116. sA63n. asteroids, meteoroids, and comets are different things

117.	sA23n.	stars emit many colors of light
118.	sA66.	optical telescopes are the only “eyes” astronomers have on the universe
119.	sA37n.	many stars are in binary systems rather than isolated by themselves
120.	sA22n.	stars move — they are not fixed on the celestial sphere
121.	sA158.	the Sun is directly overhead everywhere on Earth at noon
122.	sA142n.	meteorites continue to fall onto the Earth
123.	sA170.	Mars is the sister planet to earth in physical properties and dimensions
124.	sA186.	the Sun’s surface temperature is millions of degrees Fahrenheit
125.	sA227n.	we can see only some of the stars in the Milky Way
126.	sA9.	stars have spokes
127.	sA145n.	planes cannot fly in space
128.	sA197n.	the sun has an atmosphere
129.	sA5.	the constellations are only the stars we connect to make patterns
130.	sA182n.	the Sun causes part of the tides
131.	sA237n.	black holes have mass
132.	sA270n.	larger telescopes enable astronomers to see smaller details
133.	sA166.	Mars is the second largest planet
134.	sA222n.	the Sun is far away from the center of the Milky Way galaxy
135.	sA152.	the Earth is the only planet with an atmosphere
136.	sA58.	Mercury (closest planet to the Sun) is hot everywhere on its surface
137.	sA96n.	because the Moon scatters sunlight, the Moon does not have a mirror-like surface
138.	sA206.	the entire Sun is molten lava
139.	sA18.	the galaxy, solar system and universe are the same things

140. sA44. Pluto is always farther from the Sun than is Neptune
141. sA114n. Earth and Venus have very different atmospheres
142. sA133. the sun orbits the Earth
143. sA115. Earth is at the center of the universe
144. sA28. a nova is the most powerful explosion
145. sA252n. astronomy and astrology are different things
146. sA34. stars are fixed in space
147. sA69n. only a few planets have been known for hundreds of years
148. sA262n. the universe as a whole is changing
149. sA90n. the Moon lacks a helium atmosphere
150. sA13. all stars are evenly distributed on the celestial sphere
151. sA148. seasons are caused by speeding up and slowing down of Earth's rotation
152. sA127. auroras are caused by sunlight reflecting off polar caps
153. sA169. Mars is Hot because it is red ... Mars — god of fire
154. sA104. the side of the moon we don't see is forever "dark"
155. sA208. the Sun will explode as a nova
156. sA211n. the Sun has no chance of exploding in the "near future"
157. sA160. the Earth is flat
158. sA213n. the Sun rotates
159. sA72. there are many galaxies in a solar system
160. sA245. black holes get bigger forever and nothing can stop them from doing so
161. sA52n. planets do not orbit exactly in the plane of the ecliptic
162. sA190n. some objects in the universe are brighter than the Sun
163. sA146. a day is exactly 24 hours long
164. sA130. the air is a blue gas

165. sA108n. Venus is similar to earth in size
166. sA180. the Sun is the hottest thing in the galaxy
167. sA113. once ozone is gone from the Earth's atmosphere, it will not be replaced
168. sA79. we see all sides of the Moon each month
169. sA164n. Mars currently lacks running water on its surface
170. sA56. comets last forever
171. sA193. the Sun is a "heat planet"
172. sA267n. there is no center to the universe
173. sA16. pulsars are pulsating stars
174. sA234n. black holes are real objects in space
175. sA215. Sunspots are where meteors crash into the Sun
176. sA75n. comets are made of solid rock and ice
177. sA93. the Moon sets during daylight hours and is not visible then
178. sA150. the Earth is in the middle of the Milky Way galaxy
179. sA220. all galaxies are spiral
180. sA42. comet tails are burning — because the comet is moving so fast
181. sA3n. stars are farther away from the Earth than the Moon
182. sA83n. the Moon's distance from Earth changes
183. sA167. life, when it did exist on Mars, was quite advanced
184. sA97. the Moon will someday crash into Earth
185. sA238. black holes are like huge vacuum cleaners, sucking things in
186. sA224. the Sun is at the center of the universe
187. sA159. tides are caused just by ocean winds
188. sA116n. other planets are larger than the Earth
189. sA214. the Sun is the only source of light in the galaxy — Sunlight reflects off planets and stars so we can see them.

190. sA253. gravity will eventually pull all the planets together
191. sA45. a shooting star is actually a star whizzing across the universe or falling through the sky
192. sA135n. solar eclipses happen frequently but are not seen from all places on Earth
193. sA149n. the speed of Earth in its orbit changes as it goes around the sun
194. sA198. the Sun is the largest star
195. sA6. we are looking at stars as they are now
196. sA271. the most important function of a telescope is magnification
197. sA172. Saturn's rings are solid
198. sA59n. the day on each planet is different (not all 24 hours long)
199. sA183n. sunspots are cold spots on the Sun's surface
200. sA112. summer is warmer because we are closer to the sun during the summertime
201. sA54. all constellations look like things they are named for
202. sA233n. black holes have finite lifetimes
203. sA219n. the solar system is inside the Milky Way galaxy
204. sA68n. we have telescopes in space
205. sA25. there are exactly 12 constellations
206. sA78n. the Moon causes part of the tides
207. sA192. the Sun is made of fire
208. sA129. the day has always been 24 hours long
209. sA247. a single black hole will eventually suck in all the matter in the universe
210. sA15. all stars have same color and size
211. sA103. the Moon is larger at the horizon than when it is overhead
212. sA92n. the Moon has a rough surface

213. sA178. the Sun will burn forever

214. sA2. there are 12 zodiac constellations

215. sA41. Mercury is so named because there is much mercury on it

APPENDIX C

THE CHILDHOOD INTEREST IN ASTRONOMY SURVEY

Possible response options for each question: “never,” “occasionally,” “very often”

1. Prior to college, how often did you read astronomy books on your own?
2. Prior to college, how often did you watch astronomy programs on your own?
3. Prior to college, how often did you use binoculars or a telescope to view the night sky?
4. Prior to college, how often did you participate in an astronomy club out of pure interest?
5. Prior to college, how often did you go to planetarium shows out of pure interest?
6. Prior to college, how often did you go to an observatory out of pure interest?
7. Prior to college, how often did you keep up with news stories related to astronomical events?
8. Prior to college, how often did you choose to talk to others about astronomy, out of pure interest?

APPENDIX D

STRESS SURVEY

The questions in this scale ask you about your feelings and thoughts during the last month. In each case, you will be asked to indicate how often you felt or thought a certain way. Although some of the questions are similar, there are differences between them, so you should treat each question separately. The best approach is to answer each question fairly quickly. That is, don’t try to count up the number of times you felt a particular way; instead, indicate the alternative that seems like a reasonable estimate.

For each question, choose from the following alternatives:

0	1	2	3	4
Never	Almost Never	Sometimes	Fairly Often	Very Often

1. In the last month, how often have you felt that you were unable to control the important things in your life?
2. In the last month, how often have you felt confident about your ability to handle your personal problems?
3. In the last month, how often have you felt that things were going your way?
4. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

APPENDIX E

BLACK HOLES AND GALAXIES PRETEST

1. How are black holes created today?

(A) They form spontaneously (that is, without the need for any external matter or energy) in empty space

(B) While black holes once formed, they do not form today

(C) They form deep inside some old stars

(D) They form as the result of collisions between high-energy particles (atoms or molecules) in space

(E) There is no evidence that black holes exist, hence I don't believe they form today.

2. What is the fate of black holes?

(A) They last forever, each with a fixed mass

(B) They slowly lose mass, that is, they evaporate

(C) They continue to gain mass indefinitely

(D) They continue to grow, but one is growing faster than the others. It will eventually swallow the universe

(E) Black holes don't exist, hence the question is irrelevant

3. Black holes consist of:

(A) nothing - they are just holes in space

(B) a uniform "sea of energy"

(C) ultra-condensed concentrations of matter and energy

(D) ultra-condensed concentrations of matter, only

(E) misleading question since black holes don't really exist

4. As detectable from our universe, what shape does a black hole have?

(A) a pinhole in space

(B) a sphere

(C) a wormhole

(D) many different possible shapes, including spherical, disc, and donut shapes

(E) misleading question since black holes don't really exist

5. A spacecraft is put into circular orbit around a distant black hole. Which one of the following outcomes will happen to the spacecraft?

(A) It will be sucked straight into the black hole, like dust into a vacuum cleaner

(B) It will spiral inward, eventually falling into the black hole

(C) It will stop and hover above the black hole

(D) It will orbit around the black hole forever in the same orbit

(E) The question assumes black holes exist, which I believe they don't.

6. What would happen to an asteroid that passed into a black hole?

(A) It would be crushed and would remain in the black hole

(B) It would enter another dimension

(C) It would tunnel through a wormhole back into our universe

(D) It would remain intact in the black hole

(E) The question assumes black holes exist, which I believe they don't.

7. How do astronomers detect black holes?

(A) Astronomers detect black holes by observing their glow

(B) Astronomers detect black holes by seeing matter being ejected from them

(C) Astronomers detect black holes by putting spacecraft in orbit around them

(D) Astronomers detect black holes by their effects on nearby stars

(E) No one has actually found a black hole

8. If we boarded a spaceship to journey into a black hole, what would happen to us?

(A) We would begin to spin faster and faster

(B) We would get stuck inside the black hole

(C) We would have the ability to travel freely in time

(D) We would teleport to another dimension

(E) We would never find a black hole since they don't exist

9. How many galaxies exist in the universe today?

(A) There are no galaxies

(B) There is just our galaxy

(C) There are a few galaxies

(D) There are several thousand

(E) There are billions of galaxies

10. Relative to the Milky Way galaxy, the solar system is located:

(A) above the Milky Way

(B) between two spiral arms

(C) in a spiral arm

(D) in or near its center

(E) far away from Milky Way

11. How many distinct shapes do galaxies have?

(A) All galaxies are shaped like spirals

(B) Galaxies can be either spiral or elliptical

(C) Galaxies can have any shape other than spiral or elliptical

(D) Galaxies can be spiral, elliptical, or irregularly shaped

(E) Galaxies can have any shape

12. Where in the universe is the Milky Way located today?

(A) The Milky Way has no special location in the universe

(B) The Milky Way is located at or near the center of the universe

(C) The Milky Way *is* the universe

(D) The Milky Way is located outside the universe

(E) The Milky Way is located inside the solar system

13. Where is the Sun located relative to the Milky Way?

(A) The Sun is located above the plane of the Milky Way

(B) The Sun is located between two spiral arms of the Milky Way

(C) The Sun is located at the Milky Way's center

(D) The Sun is located in a spiral arm of the Milky Way

(E) The Sun is located far outside of the Milky Way

14. How are galaxies distributed throughout the universe?

(A) The galaxies are distributed randomly

(B) The galaxies are grouped together by size

(C) The galaxies are in large groups separated by voids

(D) The galaxies are grouped together by type or design

(E) There is only one galaxy in the universe

15. Which statement about observing stars in the Milky Way is most accurate?

(A) Astronomers can observe up to a few hundred stars in the Milky Way

(B) Astronomers can observe those stars not obscured by galactic dust

(C) Astronomers can observe hundreds of thousands of stars in the Milky Way

(D) Astronomers can observe all of the stars in the Milky Way

(E) Astronomers can observe only the brightest stars (O, B, A)

16. Which one statement about galaxy properties is correct?

(A) All galaxies are held together by gravity

(B) All galaxies have the same shape

(C) All galaxies have the same size

(D) All galaxies have the same color

(E) All galaxies rotate at the same speed

17. Which statement most accurately describes the contents of the Milky Way?

(A) The Milky Way lacks gas, but contains dust and stars

(B) The Milky Way lacks dust, but contains gas and stars

(C) The Milky Way lacks gas and dust, but contains stars

(D) The Milky Way contains gas, dust, and stars

(E) The Milky Way lacks stars, gas, and dust altogether

18. Which one of the following is true about the Milky Way?

(A) The Milky Way is still producing both stars and planets

(B) The Milky Way is still producing stars, but not planets

(C) The Milky Way is still producing planets, but not stars

(D) The Milky Way is no longer producing stars or planets

(E) The Milky Way never created its own stars or planets

APPENDIX F

FALL 2013 PRELIM QUESTIONS AND ASSOCIATED STATEMENTS FROM THE ASTRONOMY BELIEFS INVENTORY

sA2. there are 12 zodiac constellations

Prelim 1 Question 7: The astronomical zodiac consists of how many constellations?

(A) 9

(B) 12

(C) 13

(D) 52

(E) 88

sA50n. few objects orbit the Sun in circular orbits

Prelim 1 Question 14: Which of the following most accurately describes the shape of Neptune's orbit around the Sun?

(A) Circle

(B) Parabola

(C) Hyperbola

(D) Pentagon

(E) Ellipse

sA47. the asteroid belt is between Earth and Mars

Prelim 1 Question 37: Where is the asteroid belt?

(A) between Venus and Earth

(B) between Earth and Mars

(C) between Mars and Jupiter

(D) between Jupiter and Saturn

(E) between Saturn and Uranus

sA80. craters are volcanic in origin

Prelim 1 Question 38: Which statement about the surface of the Moon is correct?

(A) the highlands are older than the maria

(B) the craters on the Moon were mostly created by volcanic activity

(C) the surface is too powdery for spacecraft to land upon it

(D) there are cracks in the surface created by ice escaping from the interior

(E) part of the surface is molten today

sA164n. Mars currently lacks running water on its surface

Prelim 2 Question 2: Which statement about water on Mars is correct?

(A) Water is found in oceans there

(B) Water is found only in streams and small lakes there

(C) Only remnants of water flow have been seen on Mars' surface

(D) Water geysers have been observed on Mars

(E) No indications have been observed that water ever existed on Mars

sA50n. few objects orbit the Sun in circular orbits

Prelim 2 Question 25: What shape is the orbit of Halley's comet?

(A) circular

(B) parabolic

(C) hyperbolic

(D) elliptical

(E) triangular

sA172. Saturn's rings are solid

Prelim 2 Question 30: Saturn's rings are best described by which of the following?

(A) They are a single solid ribbon of rock and metal

(B) They are 3 or 4 solid ribbons

(C) They are a single solid ribbon of ice

(D) They are 6 or 7 rings of broken up debris

(E) They are thousands of ringlets of broken up debris

sA108n. Venus is similar to earth in size

sA170. Mars is the sister planet to earth in physical properties and dimensions

Prelim 2 Question 38: Which planet is most similar in size and composition to Earth?

(A) Mars

(B) Uranus

(C) Mercury

(D) Venus

(E) Jupiter

sA76. Jupiter's great red spot is a volcano erupting

Prelim 2 Question 39: Jupiter's Great Red Spot is most accurately called a(n)?

(A) volcano.

(B) storm.

(C) ocean.

(D) asteroid impact site.

(E) planetary ring.

sA46n. Jovian planets (Jupiter, Saturn, Uranus, Neptune) have gaseous or icy surfaces

Prelim 2 Question 41: Which statement most accurately describes descending into Jupiter?

(A) you splash into a liquid surface

(B) you land on a solid rock continent about 100 km below the clouds

(C) you land on molten rocky surface about 100 km below the clouds

(D) you descend through denser and denser gas without splashing onto a surface within 100 km of the clouds

(E) you break through a rocky crust into a liquid hydrogen/liquid helium interior

sA150. the Earth is in the middle of the Milky Way galaxy

sA222n. the Sun is far away from the center of the Milky Way galaxy

Prelim 3 Question 7: What is located at the center of the Milky Way galaxy?

(A) the solar system

(B) a supermassive black hole

(C) a pulsar

(D) the Andromeda galaxy

(E) Orono

sA184. the Sun will blow up, become a black hole, and swallow the earth

Prelim 3 Question 23: What is the final remnant of a $1\ M_\odot$ main sequence star?

(A) white dwarf

(B) black hole

(C) neutron star

(D) red dwarf

(E) red giant star

APPENDIX G

FINAL EXAM QUESTIONS AND ASSOCIATED INVENTORY STATEMENTS

sA85n. the Moon rotates even though we see only one side of it

Final Exam Question 2: Which statement about the Moon's rotation and/or revolution is correct?

(A) the Moon does not rotate

(B) the Moon rotates at the same rate that it revolves around the Earth

(C) the Moon revolves around the Earth at the same rate that the Earth rotates

(D) the Moon has retrograde rotation around the Earth

(E) the Moon rotates faster than the Earth

sA111n. Earth's axis is tilted compared to the ecliptic

Final Exam Question 3: Where is the ecliptic as seen from the Earth?

(A) directly over the Earth's north pole

(B) directly over the Earth's south pole

(C) sometimes over the Earth's north pole and sometimes over the south pole

(D) directly over the Earth's equator

(E) tilted at a 23$^1/_2$° angle compared to the celestial equator

sA78n. the Moon causes part of the tides

sA182n. the Sun causes part of the tides

Final Exam Question 4: What two bodies are the major causes of tides on Earth?

(A) Sun and Jupiter

(B) Sun and Mars

(C) Jupiter and Mars

(D) Moon and Jupiter

(E) Moon and Sun

sA2. there are 12 zodiac constellations

Final Exam Question 9: How many *zodiac* constellations are there?

(A) 12

(B) 13

(C) 25

(D) 88

(E) 110

sA185. the Sunspot cycle is 11 years long

Final Exam Question 10: What is the length of the Sun's entire magnetic cycle?

(A) about 1 month

(B) about 1 year

(C) about 11 years

(D) about 22 years

(E) about 10 billion years

sA164n. Mars currently lacks running water on its surface

Final Exam Question 19: Which of the following is found on Mars?

(A) oceans of water

(B) oceans of liquid carbon dioxide

(C) active volcanoes

(D) water flowing in rivers

(E) water ice

sA46n. Jovian planets (Jupiter, Saturn, Uranus, Neptune) have gaseous or icy surfaces

Final Exam Question 21: Which of the following best describes the surface of Jupiter?

(A) everywhere mountainous

(B) completely covered with liquid hydrogen

(C) continents and water oceans

(D) continents and liquid hydrogen oceans

(E) completely covered with molten rock (lava)

sA172. Saturn's rings are solid

Final Exam Question 22: Which statement about Saturn's rings is most accurate?

(A) Saturn's rings are solid ribbons

(B) Saturn's rings are typically composed of centimeter to meter-sized pieces

(C) Saturn's rings are uniformly smooth all around

(D) Saturn has only about one half dozen rings

(E) Saturn's rings are several thousand kilometers thick

sA171. Saturn is the only planet with rings

Final Exam Question 23: Which of the following planets does not have rings?

(A) Jupiter

(B) Uranus

(C) Neptune

(D) Venus

(E) Saturn

sA8. the north star is the brightest star in the sky

Final Exam Question 24: What is the brightest star in the night sky?

(A) Polaris (the North Star)

(B) Sirius

(C) Aldebaran

(D) Vega

(E) Arcturus

sA150. the Earth is in the middle of the Milky Way galaxy

sA222n. the Sun is far away from the center of the Milky Way galaxy

Final Exam Question 37: Where is the solar system located in the Milky Way?

(A) in the halo

(B) in the nuclear bulge

(C) in a spiral arm

(D) between two spiral arms

(E) in the nucleus

sA248. cosmic rays are light rays

Final Exam Question 40: Cosmic rays are best described by which of the following?

(A) gamma rays in space

(B) x-rays in space

(C) ultraviolet radiation in space

(D) high speed particles in space

(E) radio waves in space created by the Big Bang

sA108n. Venus is similar to earth in size

sA170. Mars is the sister planet to earth in physical properties and dimensions

Final Exam Question 44: Which planet is most similar in physical dimensions and composition to the Earth?

(A) Venus

(B) Neptune

(C) Uranus

(D) Mars

(E) Mercury

sA208. the Sun will explode as a nova

Final Exam Question 51: The Sun will end its “life” as a(n):

(A) supernova.

(B) nova.

(C) accretion disk.

(D) planetary nebula.

(E) plage.

sA50n. few objects orbit the Sun in circular orbits

Final Exam Question 67: Which of the following most accurately describes the shape of Venus’s orbit around the Sun?

(A) circular

(B) parabolic

(C) hyperbolic

(D) elliptical

(E) triangular

sA68n. we have telescopes in space

Final Exam Question 70: Where are most x-ray telescopes operated?

(A) under water

(B) on desert islands

(C) on mountain tops

(D) in jet planes

(E) in space

sA262n. the universe as a whole is changing

Final Exam Question 98: Which of the following best describes the overall motion of the universe?

(A) contracting

(B) sometimes expanding and sometimes shrinking

(C) expanding faster and faster

(D) neither expanding nor contracting

(E) astronomers do not have enough observational evidence to make any statements on the motion of the universe.

APPENDIX H

OVERLAPPING PRELIM AND FINAL EXAM QUESTIONS

- sA2. there are 12 zodiac constellations
- Prelim 1 Question 7: The astronomical zodiac consists of how many constellations?
- Final Exam Question 9: How many *zodiac* constellations are there?

- sA46n. Jovian planets (Jupiter, Saturn, Uranus, Neptune) have gaseous or icy surfaces
- Prelim 2 Question 41: Which statement most accurately describes descending into Jupiter?
- Final Exam Question 21: Which of the following best describes the surface of Jupiter?

- sA50n. few objects orbit the Sun in circular orbits
- Prelim 1 Question 14: Which of the following most accurately describes the shape of Neptune's orbit around the Sun?
- Prelim 2 Question 25: What shape is the orbit of Halley's comet?
- Final Exam Question 67: Which of the following most accurately describes the shape of Venus's orbit around the Sun?

- sA108n. Venus is similar to earth in size
- sA170. Mars is the sister planet to earth in physical properties and dimensions
- Prelim 2 Question 38: Which planet is most similar in size and composition to Earth?
- Final Exam Question 44: Which planet is most similar in physical dimensions and composition to the Earth?

- sA150. the Earth is in the middle of the Milky Way galaxy
- sA222n. the Sun is far away from the center of the Milky Way galaxy
- Prelim 3 Question 7: What is located at the center of the Milky Way galaxy?
- Final Exam Question 37: Where is the solar system located in the Milky Way?

- sA164n. Mars currently lacks running water on its surface
- Prelim 2 Question 2: Which statement about water on Mars is correct?
- Final Exam Question 19: Which of the following is found on Mars?

- sA172. Saturn's rings are solid
- Prelim 2 Question 30: Saturn's rings are best described by which of the following?
- Final Exam Question 22: Which statement about Saturn's rings is most accurate?

APPENDIX I

OVERLAPPING ATTENDANCE QUESTIONS WITH ITEMS FROM THE ASTRONOMY BELIEFS INVENTORY

Attendance Question 1

n = 171, no inter-rater analysis conducted

sA2. there are 12 zodiac constellations

September 5, 2013 (immediately at the start of class, prior to lecture). How many zodiac constellations are there?

1. (36%) 13
2. (34%) 12
3. (35%) Other, vague, unsure

Attendance Question 2

n = 162, κ = 0.74

sA112. summer is warmer because we are closer to the sun during the summertime

sA148. seasons are caused by speeding up and slowing down of Earth's rotation

September 5, 2013 (at the end of class, as normal). What causes the seasons?)

1. (38%) Tilt of the Earth
2. (17%) Changing distance between the Earth and the Sun
3. (46%) All other (less logical) (e.g. "the Earth's rotation")

Attendance Question 3

n = 166, no inter-rater analysis conducted

sA50n. few objects orbit the Sun in circular orbits

September 10, 2013. What is the shape of the Earth's orbit around the sun?

1. (86%) elliptical or oval
2. (13%) circular or sphere
3. (2%) other

Attendance Question 4

n = 154, κ = 0.62

sA161. Mars is green (from plant life)

September 30, 2013. Have astronomers found life on Mars?

1. (38%) No, (5%) Only as fossils (traces)
2. (18%) Other (e.g. found water/ice, didn't answer question)
3. (40%) Yes

Attendance Question 5

n = 150, no inter-rater analysis conducted

sA46n. Jovian planets (Jupiter, Saturn, Uranus, Neptune) have gaseous or icy surfaces

October 3, 2013. Describe the surface of Jupiter

1. (37%) Gas, liquid (but not water), or "no surface" - but not solid
2. (35%) Solid - not gas (e.g. "cratered")
3. (29%) All other, including mixed surface types (e.g. gas and rocky)

Attendance Question 6

n = 138, κ = 0.71

sA172. Saturn's rings are solid

October 10, 2013. Briefly describe the rings of Saturn

1. (78%) Particles, ice, debris, rocks, asteroids, metals, “like Jupiter’s”
2. (14%) Solid continuous matter, inc. gravity-held “material” or “rocky elements”
3. (8%) Gas only, clouds, or liquid

Attendance Question 7

$n = 120$, no inter-rater analysis conducted

sA189. the Sun is the brightest star in universe

sA190n. some objects in the universe are brighter than the Sun

November 7, 2013. How bright is the sun compared to other stars?

1. (40%) average or about average
2. (37%) dimmer, or (22%) brighter
3. (2%) the Sun is the brightest star

Attendance Question 8

$n = 128$, $\kappa = 1.00$

sA248. cosmic rays are light rays

November 12, 2013. What are cosmic rays?

1. (1%) High-energy particles from space, (2%) High-energy particles (not from space or unspecified)
2. (21%) Light or light rays
3. (77%) All other (inc. vague, unspecified, multiple responses) (e.g. “rays coming from stars,” “radiation,” “shock waves from supernovas”)

Attendance Question 9

n = 125, no inter-rater analysis conducted

sA150. the Earth is in the middle of the Milky Way galaxy

sA222n. the Sun is far away from the center of the Milky Way galaxy

November 21, 2013. Where in the Milky Way are we?

1. (2.4%) Between two spiral arms, (30%) On or in a spiral arm
2. (30%) In or near the center
3. (37%) All other or too vague to tell (e.g. “upper left,” “off to the side,” “on the edge”)

APPENDIX J

VIDEO LECTURE PRETEST QUESTIONS

J.1 Black Holes

1. How are black holes created today?

(A) They form spontaneously (that is, without the need for any external matter or energy) in empty space

(B) While black holes once formed, they do not form today

(C) They form deep inside some old stars

(D) They form as the result of collisions between high-energy particles (atoms or molecules) in space

(E) There is no evidence that black holes exist, hence I don't believe they form today.

2. What is the fate of black holes?

(A) They last forever, each with a fixed mass

(B) They slowly lose mass, that is, they evaporate

(C) They continue to gain mass indefinitely

(D) They continue to grow, but one is growing faster than the others. It will eventually swallow the universe

(E) Black holes don't exist, hence the question is irrelevant

3. Black holes consist of:

(A) nothing - they are just holes in space

(B) a uniform “sea of energy”

(C) ultra-condensed concentrations of matter and energy

(D) ultra-condensed concentrations of matter, only

(E) misleading question since black holes don’t really exist

4. As detectable from our universe, what shape does a black hole have?

(A) a pinhole in space

(B) a sphere

(C) a wormhole

(D) many different possible shapes, including spherical, disc, and donut shapes

(E) misleading question since black holes don’t really exist

5. A spacecraft is put into circular orbit around a distant black hole. Which one of the following outcomes will happen to the spacecraft?

(A) It will be sucked straight into the black hole, like dust into a vacuum cleaner

(B) It will spiral inward, eventually falling into the black hole

(C) It will stop and hover above the black hole

(D) It will orbit around the black hole forever in the same orbit

(E) The question assumes black holes exist, which I believe they don’t.

6. What would happen to an asteroid that passed into a black hole?

(A) It would be crushed and would remain in the black hole

(B) It would enter another dimension

(C) It would tunnel through a wormhole back into our universe

(D) It would remain intact in the black hole

(E) The question assumes black holes exist, which I believe they don't.

7. How do astronomers detect black holes?

(A) Astronomers detect black holes by observing their glow

(B) Astronomers detect black holes by seeing matter being ejected from them

(C) Astronomers detect black holes by putting spacecraft in orbit around them

(D) Astronomers detect black holes by their effects on nearby stars

(E) No one has actually found a black hole

8. If we boarded a spaceship to journey into a black hole, what would happen to us?

(A) We would begin to spin faster and faster

(B) We would get stuck inside the black hole

(C) We would have the ability to travel freely in time

(D) We would teleport to another dimension

(E) We would never find a black hole since they don't exist

J.2 Galaxies

1. How many galaxies exist in the universe today?

(A) There are no galaxies

(B) There is just our galaxy

(C) There are a few galaxies

(D) There are several thousand

(E) There are billions of galaxies

2. Relative to the Milky Way galaxy, the solar system is located:

(A) above the Milky Way

(B) between two spiral arms

(C) in a spiral arm

(D) in or near its center

(E) far away from Milky Way

3. How many distinct shapes do galaxies have?

(A) All galaxies are shaped like spirals

(B) Galaxies can be either spiral or elliptical

(C) Galaxies can have any shape other than spiral or elliptical

(D) Galaxies can be spiral, elliptical, or irregularly shaped

(E) Galaxies can have any shape

4. Where in the universe is the Milky Way located today?

(A) The Milky Way has no special location in the universe

(B) The Milky Way is located at or near the center of the universe

(C) The Milky Way *is* the universe

(D) The Milky Way is located outside the universe

(E) The Milky Way is located inside the solar system

5. Where is the Sun located relative to the Milky Way?

(A) The Sun is located above the plane of the Milky Way

(B) The Sun is located between two spiral arms of the Milky Way

(C) The Sun is located at the Milky Way's center

(D) The Sun is located in a spiral arm of the Milky Way

(E) The Sun is located far outside of the Milky Way

6. How are galaxies distributed throughout the universe?

(A) The galaxies are distributed randomly

(B) The galaxies are grouped together by size

(C) The galaxies are in large groups separated by voids

(D) The galaxies are grouped together by type or design

(E) There is only one galaxy in the universe

7. Which statement about observing stars in the Milky Way is most accurate?

(A) Astronomers can observe up to a few hundred stars in the Milky Way

(B) Astronomers can observe those stars not obscured by galactic dust

(C) Astronomers can observe *hundreds of thousands* of stars in the Milky Way

(D) Astronomers can observe all of the stars in the Milky Way

(E) Astronomers can observe only the brightest stars (O, B, A)

8. Which one statement about galaxy properties is correct?

(A) All galaxies are held together by gravity

(B) All galaxies have the same shape

(C) All galaxies have the same size

(D) All galaxies have the same color

(E) All galaxies rotate at the same speed

9. Which statement most accurately describes the contents of the Milky Way?

(A) The Milky Way lacks gas, but contains dust and stars

(B) The Milky Way lacks dust, but contains gas and stars

(C) The Milky Way lacks stars, but contains gas and dust

(D) The Milky Way lacks gas and dust, but contains stars

(E) The Milky Way contains gas, dust, and stars

10. Which one of the following is true about planets and stars in the Milky Way?

(A) The Milky Way is still producing both stars and planets

(B) The Milky Way is still producing stars, but not planets

(C) The Milky Way is still producing planets, but not stars

(D) The Milky Way is no longer producing stars or planets

(E) The Milky Way never created its own stars or planets

APPENDIX K

OPTIMAL ORDERS TO TEACH CONCEPTS IN EACH TOPIC OF THE ASTRONOMY BELIEFS INVENTORY

K.1 Stars

Group	Order of Statements
Children and Adolescents	
2	sA11, sA18, sA6, sA17, sA7
3	sA32, sA15, sA14, sA3, sA30
1	sA33, sA34, sA20, sA23, sA22
5	sA1, sA9, sA10, sA13
7	sA4, sA8, sA12, sA16
6	sA24, sA29, sA27
8	sA5, sA25, sA2
4	sA37, sA28, sA21, sA31, sA35
Adults	
3	sA32, sA15, sA14, sA30, sA3
5	sA1, sA10, sA13, sA9
2	sA18, sA11, sA17, sA7, sA6
1	sA20, sA33, sA23, sA34, sA22
7	sA4, sA8, sA16, sA12
8	sA25, sA2, sA5
4	sA37, sA28, sA35, sA21, sA31
6	sA24, sA29, sA27

Table K.1. Optimal orders to teach statements about the stars

K.2 Solar system

Group	Order of Statements
	Children and Adolescents
8	sA59, sA57
2	sA48, sA67, sA68, sA56
5	sA51, sA43, sA53
7	sA63, sA66, sA62
4	sA54, sA46, sA44, sA58
1	sA45, sA41, sA42, sA60, sA40
9	sA69, sA52
3	sA50, sA76, sA47, sA72, sA49
6	sA75, sA70
	Adults
1	sA60, sA41, sA40, sA42, sA45
8	sA59, sA57
5	sA51, sA43, sA53
7	sA62, sA63, sA66
2	sA56, sA48, sA68, sA67
3	sA50, sA76, sA72, sA49, sA47
9	sA52, sA69
4	sA46, sA58, sA54, sA44
6	sA70, sA75

Table K.2. Optimal orders to teach statements about the solar system

K.3 Moon

Group	Order of Statements
	Children and Adolescents
4	sA77, sA93, sA102, sA94, sA96
3	sA92, sA78, sA84, sA79
1	sA91, sA89, sA87, sA80, sA88, sA105, sA90
5	sA98, sA97
2	sA85, sA83, sA103, sA104, sA100, sA99
	Adults
4	sA94, sA77, sA102, sA93, sA96
3	sA79, sA92, sA84, sA78
1	sA88, sA89, sA87, sA91, sA90, sA105, sA80
5	sA97, sA98
2	sA85, sA104, sA83, sA100, sA99, sA103

Table K.3. Optimal orders to teach statements about the Moon

K.4 Venus, Mars, Saturn

Group	Order of Statements
	Children and Adolescents
2	sA171, sA172, sA176
1	sA169, sA167, sA168, sA161, sA166
3	sA109, sA107, sA106
5	sA164, sA165, sA174
4	sA108, sA170, sA110
	Adults
1	sA169, sA168, sA161, sA166, sA167
2	sA176, sA171, sA172
3	sA106, sA107, sA109
5	sA164, sA175, sA165
4	sA110, sA108, sA170

Table K.4. Optimal orders to teach statements about Venus, Mars, and Saturn

K.5 Earth

Group	Order of Statements
	Children and Adolescents
2	sA116, sA133, sA115, sA145
4	sA147, sA146, sA129
9	sA112, sA111
1	sA143, sA152, sA137, sA159, sA158, sA128, sA156, sA122
5	sA160, sA141, sA142, sA148, sA157
3	sA154, sA135, sA126, sA114, sA113
7	sA151, sA127, sA125
6	sA144, sA118, sA149
8	sA130, sA150, sA153, sA131
	Adults
5	sA160, sA157, sA142, sA148, sA141
1	sA143, sA159, sA128, sA152, sA156, sA137, sA122, sA158
2	sA145, sA116, sA133, sA115
8	sA130, sA150, sA153, sA131
9	sA111, sA112
7	sA125, sA151, sA127
3	sA135, sA114, sA154, sA126, sA113
4	sA147, sA146, sA129
6	sA144, sA118, sA149

Table K.5. Optimal orders to teach statements about the Earth

K.6 Sun

Group	Order of Statements
	Children and Adolescents
2	sA192, sA178, sA206, sA181, sA182
1	sA180, sA198, sA189, sA177, sA209, sA190
4	sA193, sA214, sA211, sA196, sA215
3	sA204, sA199, sA200, sA187
8	sA188, sA186
5	sA185, sA184, sA183, sA191, sA213
6	sA217, sA197, sA201, sA208
7	sA202
	Adults
7	sA202
4	sA215, sA196, sA193, sA214, sA211
2	sA206, sA178, sA192, sA182, sA181
3	sA204, sA187, sA199, sA200
1	sA177, sA209, sA189, sA198, sA190, sA180
8	sA188, sA186
5	sA184, sA213, sA183, sA185, sA191
6	sA197, sA201, sA217, sA208

Table K.6. Optimal orders to teach statements about the Sun

K.7 Galaxies

Group	Order of Statements
	Children and Adolescents
2	sA224, sA222, sA221
1	sA218, sA225, sA228, sA230, sA231, sA219, sA220
3	sA227, sA226
	Adults
1	sA218, sA219, sA228, sA220, sA231, sA230, sA225
2	sA224, sA221, sA222
3	sA227, sA226

Table K.7. Optimal orders to teach galaxy concepts

K.8 Black holes

Group	Order of Statements
	Children and Adolescents
3	sA240, sA244, sA243
2	sA246, sA247, sA245, sA242, sA233
1	sA235, sA232, sA234, sA237 sA238
	Adults
3	sA243, sA240, sA244
2	sA242, sA247, sA246, sA245, sA233
1	sA234, sA232, sA237, sA235, sA238

Table K.8. Optimal orders to teach black hole concepts

K.9 General astrophysics

Group	Order of Statements
	Children and Adolescents
2	sA254, sA262, sA258, sA270
3	sA252, sA271, sA273, sA272
1	sA255, sA256, sA259, sA267, sA263
4	sA261, sA253, sA248
	Adults
2	sA258, sA270, sA240, sA254
4	sA253, sA248, sA261
3	sA252, sA271, sA273, sA272
1	sA255, sA263, sA267, sA259, sA256

Table K.9. Optimal orders to teach statements about astrophysics in general

BIOGRAPHY OF THE AUTHOR

Andrej Favia was born in Montvale, New Jersey, on July 20, 1985. As a child, he was active in learning about math and science, especially astronomy, through books and videos. He has been following his lifelong passion for astronomy ever since. He received his high school education from Pascack Hills High School in Montvale, New Jersey, in 2003. He moved to Maine and attended the University of Southern Maine in 2003. He obtained his Bachelor of Arts degree in physics in 2007 and graduated summa cum laude.

In September 2007, Andrej enrolled for graduate study in the department of Physics and Astronomy at the University of Maine and served as a Teaching Assistant in physics and astronomy labs. His current research interests include astronomy education and astrophysics. He is a member of Phi Kappa Phi. He is a candidate for the Doctor of Philosophy degree in Physics from the University of Maine in August 2014.