A DIRECT EXAMINATION OF COLLEGE STUDENT MISCONCEPTIONS IN ASTRONOMY: A NEW INSTRUMENT

Andrej Favia[1], Neil F. Comins[1], Geoffrey L. Thorpe[2], and David J. Batuski[1], *University of Maine*

Received April 22, 2014; Accepted May 21, 2014

Abstract: This is the first in a series of papers in which we examine the persistence of 215 common misconceptions in astronomy and suggest correlations among them in an effort to improve the effectiveness of astronomy instruction. Each misconception is based on a commonly-held incorrect belief by college students taking introductory astronomy. At the University of Maine, the course is taught in alternating semesters by Neil F. Comins and David J. Batuski. A total of 639 students over six semesters between 2009 and 2013 completed a survey based on these misconceptions. The survey is a new instrument in that it permits one to indicate either endorsement or rejection of each misconception at various stages in one's life. We present *two versions* of the survey: one in which all statements are presented as misconceptions, and one in which both true and false statements are presented. We test the validity of the survey data and present a preliminary analysis of the data for both versions of the survey. We show that the length of the survey and the presentation order of the statements are unlikely to affect the data. We also show that the reported degree of misconception endorsement may be affected by the phrasing of the statements, that is, whether or not the statements are all false or a mixture of true and false statements.

Keywords: students - non-science majors - general astronomy - Astro 101 - misconceptions - undergraduate education

INTRODUCTION

Since the 1980s, the process of how students learn concepts related to perceptions of motion, the colors of objects, and heat, among many other topics, has been studied, with the subjects of the studies ranging from children to adults (Anderson & Smith, 1988; diSessa, 1982; Flavell, Green, & Flavell, 1986; Kempton, 1987; Posner, Strike, & Hewson, 1982; Sadler, 1998; White, 1982; Vosniadou, 1994; Vosniadou & Brewer, 1992). These studies draw four 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, students in the class may retain any one of a multitude of inappropriate models to explain their observations of relatively simple physical phenomena. Third, misconceptions are strongly-held incorrect beliefs, so much that, as Vosniadou (1994) states, instructors are encouraged "to understand [the misconceptions] and to take them into consideration in the design of instruction" (p. 66). Fourth, these studies support a growing body of research (Clark, Kirschner, & Sweller, 2012, and references therein) showing that, despite the pedagogical efforts of a wide range of instructors in the field, misconceptions in astronomy remain persistent.

The effect of misconceptions on understanding astronomy concepts has been analyzed in a number of studies including Bailey, Prather, Johnson, and Slater (2009), Sadler et al. (2010), Vosniadou, Vamvakoussi, and

[1] Department of Physics and Astronomy
[2] Department of Psychology

Skopeliti (2008), and Wallace, Prather, and Duncan (2011). We note in particular that Bailey et al. (2009) assessed pre-instructional ideas about stars and star formation held by 2,200 non-science majors taking an introductory astronomy course ('pre-instructional' in this case 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"). 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 et al., 2008).

One approach to assessing how much students learn during instruction about a particular topic is the development of pretests and posttests (e.g. as in the aforementioned studies) to serve as probes for measuring learning. The use of a pretest immediately before instruction and a posttest immediately after instruction presents a controlled environment. In fact, the vast majority of the aforementioned studies rely on recording student responses to a predetermined set of guided questions. An example of a multiple-choice test designed with such questions in mind is presented by Sadler (1998). 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 by themselves 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, Sinatra, & Nussbaum, 2013; Prather et al., 2004). These tests provide support for conducting studies in educational research in which the data are acquired months after the pretest.

To address specifically the persistence of misconceptions, one ought to study the long-term effects of instruction. As Vosniadou (1994) reminds us, one *should* be well informed of the misconceptions. Instead of designing and implementing a multiple-choice pretest, 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. Such a design provides students with the opportunity to give real-time feedback. The design also allows the option for students to indicate approximately when, in their lives, they harbored a misconception, or still endorse it even after instruction in the course, or simply indicate if they have never heard of it before. This last option is generally not provided on multiple-choice tests.

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.

The purpose of this research is to study the misconceptions that students bring to the college astronomy classroom; the focus of our research is on students enrolled in the introductory astronomy course at the University of Maine from the Fall 2009 to Fall 2013 semesters. 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 ameliorates the misconceptions most effectively. Our 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. We begin our analysis in this paper, the first in a series, by presenting a new instrument with which we gather our data to achieve these goals.

METHOD

We developed a new survey consisting of an inventory of statements, presented as short beliefs (e.g., “all stars are white,” “Saturn's rings are solid”), spanning all topics in astronomy. From an initially larger item pool, we selected 215 misconception-based statements organized by topics; these statements are 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. When the actual inventory was administered to the students, the labels were omitted. *The list of 215 items comprises the Astronomy Beliefs Inventory* (ABI). The ABI measures the extent to which students endorse any of these beliefs. The ABI also allows the student to indicate if the student had heard of any of the beliefs prior to college. The ABI was made available to students taking the introductory-level astronomy lecture at the University of Maine, on a voluntary basis for extra credit. In sections taught by co-author Neil F. Comins (NFC), students who opted out were allowed to write an essay for equivalent extra credit. On average, students who volunteered to respond to the ABI required about one hour (a two-hour timeslot was provided). Further along, we will examine the effect of student fatigue on the reliability of the ABI data.

In this section, we outline the development and administration of *two versions* of the ABI. The ABI was administered at the end of each of six semesters at the University of Maine: Fall 2009, Fall 2010, Fall 2011, Fall 2012, Spring 2013, and Fall 2013. The total sample size for all six semesters is N = 639, of which 341 students are male, and 297 students are female. Demographic information is available for all but one student. The average age of the sample is M = 20.0, SD = 3.8 years. 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, African-American, and other/unspecified are 84.6%, 2.0%, 1.9%, 1.4%, 0.9%, and 8.8%. Instructors for the course are NFC and David J. Batuski (DJB). Table 1 presents a summary of the ABI administrations, with the formats (I and II) to be discussed shortly.

Table 1

Administrations of Misconception-Based Statement Lists, Per Semester

Semester	Instructor	Class Size	Sample Size	Statement Count	Format
Fall 2009	NFC	188	114	267	I
Fall 2010	NFC	175	107	235	I
Fall 2011	NFC	171	91	235	I
Fall 2012	NFC	170	91	235	I
Spring 2013	DJB	192	126	235	II
Fall 2013	NFC	174	110	235	II

As noted in Table 1, the course was taught by either of two instructors. Each instructor employs a slightly different teaching pedagogy. In 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). To inquire of the misconceptions held by his students, at the end of each class, NFC asks a misconception-based attendance question about a topic to be lectured in the subsequent class. That is, the question is asked before the related topic is discussed. For example, if the question is “How many zodiac constellations are there?” then the subsequent lecture would include a discussion about zodiac constellations.

In lecture, DJB teaches the material by presenting facts in a traditional manner. In following this traditional framework, DJB places less emphasis than NFC on explicitly announcing common misinformation held by the students during lecture. DJB takes attendance by the use of clicker questions, which ask the students to provide feedback on various concepts in astronomy in real time. The use of such clicker questions has been previously used in astronomy classrooms at other universities (Prather & Brissenden, 2009). The study by Prather and Brissenden promotes the use of clicker questions and claims that they improve (i) student understanding of course concepts and (ii) exam scores. 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). In the course taught by DJB, the clicker questions are not usually misconception driven, although misconceptions are frequently involved or probed.

By 2001, NFC had sufficient data on student misconceptions in all general topics pertaining to astronomy (Comins, 2001, 2014) to teach to his students in the context of the misconceptions. In 2009, NFC administered a preliminary list of 267 statements (nearly all of which are false) to his students. Many of the items in the original item pool were eliminated due to issues with clarity, while a few new statements were added; the revised inventory consisted of 235 items. Two versions of the inventory were developed. The first version consisted of either 235 or 267 false statements, depending on the semester. Students were asked to indicate (on Scantron sheets) approximately when in their lives they believed each statement, if ever, or have never heard of it before. Students were also encouraged to write a correction to any statements they did not believe. Directions for completing the first format (Format I) of the inventory are presented in Table 2. A second version of the inventory was later developed, as discussed in the subsequent paragraph.

Table 2

Directions for Completing Format I of the Inventory

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.

A special format of the inventory was used for the Spring 2013 and Fall 2013 semesters. Directions for completing the second format (Format II) of the inventory are presented in Table 3.

Table 3

Directions for Completing Format II of the Inventory

*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.

In the second format (Format II), of the 215 statements under consideration, 129 statements were phrased as false, and 86 statements were phrased as true. For the purposes of our analysis, a false statement is a statement phrased as a misconception (e.g., sA111, “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). Often incorrect statements were made “correct” simply by reversing their direction. In addition, the sequence of the statement presentation was randomized, and three different random-order forms (designed #1, #2, #3) were presented to the students. All three forms contained the same statements, just presented in different sequences.

The advantage of administering Format II is that the mixture of true-false statements eliminates much of the bias introduced from an instrument in which nearly all statements were false. By comparing the responses to the two formats, we were able to test the data of the ABI for convergent validity.

A master code was developed for all responses to the first and second formats of the inventories. The motivation for the codes is the desire to preserve the sense of "timeline," as to approximately how late in one's life does one abandon a misconception. We developed three codes, 1, 2, and 3, which indicate relative degrees of misconception *retainment*, where we 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 got rid 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 4. That students may report and then recall 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 we will show in a later paper, however, there is little concern for the accuracy of these reports, because reports on the ABI are comparable to that of instruments designed in a more traditional multiple-choice format.

Table 4

Codes for Three Relative Degrees of Misconception Retainment

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

Note that in Format II of the ABI, 86 of the 215 statements under consideration were changed from incorrect to scientifically accurate, as discussed on page 5. 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 Format II of the ABI.

TESTING THE VALIDITY OF THE ABI

Scoring the Data

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. To analyze our data, we first quantified the data by using the master codes in Table 4, then we computed the mean misconception retainment score for each student, which is the mean over the responses to all statements. To calculate the mean misconception retainment score for each student, we summed over the misconception retainment scores (1, 2, 3) for each item on the inventory, then divided the result by the number of items to which the student responded. The range of possible scores for each student is from 1 to 3, where students with scores between 2 and 3 tend to endorse misconceptions even after instruction, and students with scores between 1 and 2 tend to dispel misconceptions prior to or during instruction. Of the total sample, 89% of the students responded to all 215 statements under consideration, and 98% of the students responded to at least 212 of the statements.

To assess the relative difficulty of the statements, for each statement, we calculated the mean "score" using all of the student responses, already coded as degrees of misconception retainment (Table 4). 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 also reported in Appendix A. Statements with scores between 2 and 3 are associated with misconceptions that are relatively difficult to dispel, and statements with scores between 1 and 2 are associated with misconceptions that are relatively easy to dispel. 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.

The histogram in Figure 1 shows the distribution of the misconception retainment scores, presented in Appendix A. The distribution is a continuum from 1, which represents the lowest degree of retainment, to 3, which represents the highest degree of retainment. Misconceptions with retainment scores much closer to 2 than the extremes tend not to be readily dispelled, except through instruction. Examples of such misconceptions are "the Sun is hottest on its surface," with a retainment score of 2.02, and "sunspots are constant fixtures on the Sun," with a retainment score of 1.97. These misconceptions are so close to 2 that they almost equally likely to persist until one is instructed otherwise.

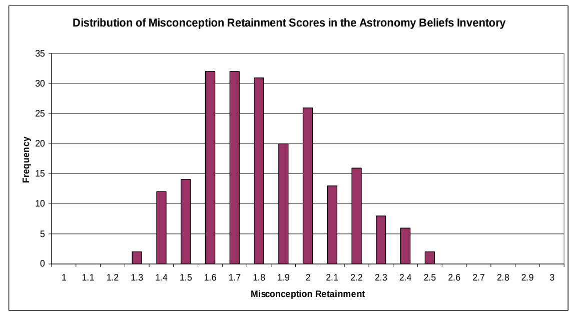

Figure 1. Distribution of misconception persistence scores ("retainment"). Average is 1.88, SD = .266.

Effect of Statement Presentation Order

We introduced Format II of the inventory (the directions of which are in Table 3) to address criticisms regarding statement presentation order. 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. If presentation order plays a significant influence on the responses to the ABI, then one form would have significantly different responses, either for a particularly topic of the ABI or for the entire instrument, than other forms.

To discern whether or not any of the three forms had significantly different scores from the others, we performed independent ANalysis Of VAriance (ANOVA) tests on the three different random orders, *for each topic of the ABI*. The scores under consideration consist of the overall degree of misconception retainment, per student, averaged over the items for each individual topic of the ABI. For our analyses, the topics of the ABI are treated independently and are *not* combined together, so that one ANOVA test is performed per topic, within each semester. According to French, Macedo, Poulsen, Waterson, and Yu (2008), ANOVA "tests for the difference in means between two or more groups" (p. 1), while Multivariate ANalysis Of VAriance (MANOVA) considers two or more *dependent* variables to quantify the difference in means. For the purpose of our analyses, each topic of the ABI is treated independently, and, for each topic, we consider only one variable, the overall degree of misconception retainment, averaged over exclusively the items within it. Therefore, we have elected to conduct ANOVA tests on our data.

We now briefly describe a few important statistical parameters regarding the ANOVA test. The Levene statistic, W, measures the deviation in the homogeneity of variances in the scores and is considered a relatively robust statistic compared to similar statistics (Borkowski, 2014; Hole, 2013). Associated with the Levene statistic is the significance p_V in the differences in the variances, which represents how often one would obtain a value of at least W for the Levene statistic by chance. The F-ratio is the ratio of the variance of the scores between groups vs. the variance of the scores within groups. Associated with the F-ratio is the usual p statistic, which explicitly determines if the means among the groups are significantly different from each other. Values of $p < .05$ are considered statistically significant and are marked with an asterisk (*). These values are significant, because they indicate that there is at least a one in 20 chance of being incorrect when drawing the conclusion that the means are significantly different. Because of the number of univariate comparisons (18 in total, with nine topics for each semester, see Table 5), there is naturally the possibility that one test result may be statistically significant by chance. To employ a less-conservative p value, one may apply a Bonferroni correction (Bland & Altman, 1995), which is an adjustment for the p value based on the number of tests. For a single univariate test, we set the cutoff at .05, whereas for K tests, a statistically significant result on a test would occur if $p < .05/K$, according to the Bonferroni correction.

An additional statistic, however, is needed 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. Note that one does not actually need a significant result to calculate the effect size. In an ANOVA test, several measures are used to calculate the effect size (Becker, 2000; Cohen, 1988), of which η^2 and ω^2 are summarized here. A common estimation of the effect size is given by η^2 equaling $SS_{Between} / SS_{Total}$ where $SS_{Between}$ is the sum of squares between individual groups, SS_{Within} is the sum of squares within the groups, and SS_{Total} is the sum of these quantities. 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 accounts for variation within groups to minimize the bias in the estimator. The revised estimator ω^2 of the effect size is given by

$$\omega^2 = \frac{SS_{Between} - df_{Between} \times MS_{Within}}{SS_{Total} + MS_{Within}} \qquad \text{where} \qquad MS_{Within} = \left(\frac{SS_{Within}}{df_{Within}}\right)$$

is the mean square of the data within groups, and *df* is the degrees of freedom. We report both η^2 and ω^2 to estimate effect sizes as appropriate. It is possible that, for some not statistically significant results, ω^2 may become negative, which essentially means that the effect size is negligibly small.

Table 5

Significance of Differences Among the Three Random Sequences of ABI Statements, with Terms as Defined in the Text

	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	.624	1.403	.250	.022	.006
Solar System	1.86	0.32	1.261	.287	3.048	.051	.047	.032
Moon	1.79	0.28	0.367	.693	3.935	.022*	.060	.044
Three Planets	1.81	0.32	0.669	.514	1.450	.239	.023	.007
Earth	1.92	0.27	0.782	.460	1.320	.271	.021	.005
Sun	2.00	0.27	0.657	.520	3.026	.052	.047	.031
Galaxies	1.83	0.32	1.994	.141	1.929	.150	.030	.015
Black Holes	2.01	0.29	0.944	.392	1.990	.141	.031	.015
General Astrophysics	2.10	0.25	0.046	.955	1.528	.221	.024	.008
All 215 Statements	**1.91**	**0.25**	**0.625**	**.537**	**2.567**	**.081**	**.040**	**.024**

	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	.898	0.036	.965	.001	-.018
Solar System	1.87	0.30	1.744	.180	0.450	.639	.008	-.010
Moon	1.81	0.29	0.353	.704	0.279	.757	.005	-.013
Three Planets	1.77	0.33	0.522	.595	0.706	.496	.013	-.005
Earth	1.88	0.26	0.306	.737	0.065	.937	.001	-.017
Sun	2.00	0.25	1.199	.305	0.244	.784	.005	.003
Galaxies	1.82	0.35	0.313	.732	0.001	.999	.000	-.019
Black Holes	2.10	0.30	0.065	.937	0.094	.910	.002	-.017
General Astrophysics	2.07	0.26	0.816	.445	0.526	.593	.001	-.009
All 215 Statements	**1.91**	**0.25**	**0.126**	**.882**	**0.102**	**.903**	**.002**	**-.017**

Having outlined the statistical variables, we are now ready to present our results. Table 5 presents 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 5 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, the significance of the difference in means among groups, and the effect sizes as estimated by η^2 and ω^2. 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 1-3 for all tests.

With regard to the data for both the Spring 2013 and Fall 2013 semesters, Table 5 shows that there were generally no statistically significant differences in the ABI scores among the three forms, indicating that the order of statement presentation had no significant influence on student responses. Only on form #1 in the Spring 2013 semester did students report having a marginally higher degree of misconception endorsement than on forms #2 and #3 for that semester. Because we are using independent univariate tests, however, we do not see the p value of .022 to be significant in the context of the Bonferroni correction.

Effect of Fatigue

Students typically spend between one and one and a half hours responding to all 235 items (or 267 items in the case of the Fall 2009 semester), raising the question that student fatigue, at some point during the response process, may sacrifice the validity of responses provided by the students thereafter. If so, then responses to the ABI ought to become less meaningful in the later sections, and this can be tested, since the correlations between earlier and later items would be low. We used data from all of the original 235 statements in the inventories administered to the students in the Spring 2013 and Fall 2013 semesters. We 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 forms contain statements in essentially random orders, the scores on the first and second halves are expected to be well correlated, except if fatigue interferes with the response process. We also report the coefficient of determination (or r^2, where r is the correlation coefficient). 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 variation in the other (Taylor, 1990). We further report the degrees of freedom, df, which is one less than the number of subjects, per semester. Individual scores range from 1-3 for all tests. For the Spring 2013 semester, where $n = 126$, we report summary statistics for the first half ($M = 1.88$, $SD = 0.26$) and the second half ($M = 1.95$, $SD = 0.26$). For the Fall 2013 semester, where $n = 110$, we analogously report summary statistics for the first half ($M = 1.87$, $SD = 0.25$) and the second half ($M = 1.96$, $SD = 0.26$). The correlations between the first and second halves are .772 ($r^2 = .59$, $df = 125$, $p < .0005$) for the Spring 2013 semester and .864 ($r^2 = .74$, $df = 109$, $p < .0005$) for the Fall 2013 semester. On the basis of this analysis, there is no evidence that students respond differently between the first and second halves of the inventory, which is consistent with the hypothesis that fatigue does not sacrifice the validity of the data.

As an additional check on the influence of fatigue (if any) on student responses to the inventory, we 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) (Schmitt, 1996), sometimes referred to as Cronbach's alpha. Coefficient α depends on the number of items in a test. Values of $\alpha \geq .70$ represent a group of items with "adequate" internal consistency. Using the original 235 statements, as administered to the students, we 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, we report summary statistics for the Earth topic ($n = 126$, $M = 1.92$, $SD = 0.27$) and the Sun topic ($n = 126$, $M = 2.00$, $SD = 0.27$); these are the same as in Table 5. In the Spring 2013 semester, values of α ranged from .79 to .84 for the Earth topic and .74 to .81 for the Sun topic. For the Fall 2013 semester, we report summary statistics for the Earth topic ($n = 110$, $M = 1.88$, $SD = 0.26$) and the Sun topic ($n = 110$, $M = 2.00$, $SD = 0.26$). In the Fall 2013 semester, α ranged from .83 to .84 for the Earth topic and .78 to .85 for the Sun topic.

We now briefly interpret the values of the coefficient α. While α may seem low given the large number of items per topic, the reader should be aware that statements within each topic of the ABI are associated with a particular factor structure that describes the inter-item correlations. In a forthcoming paper, we 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 high inter-item correlations; however, inter-item correlations between statements of *different* groups tend to be much lower. These results are consistent with α to be low for each topic as a whole, since not every item inter-correlates strongly with every other item in the topic. The same is true for all other topics in the ABI. Hence, the reader should not be alarmed by the somewhat low values of α. 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.

Effect of False vs. True Statements

For each semester, we calculated the fraction of misconceptions endorsed even after instruction in the course. To score the data, we took the ABI statement responses, previously coded as degrees of misconception retainment as described in Table 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 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.

Table 6

Mean Fraction of Incorrect Beliefs Endorsed Per Semester, Using All 215 Statements

Semester	Sample Size	Mean Fraction Believed	Std. Dev.	Format
Fall 2009	114	.200	.109	I
Fall 2010	107	.173	.109	I
Fall 2011	91	.186	.109	I
Fall 2012	91	.117	.097	I
Spring 2013	126	.275	.105	II
Fall 2013	110	.252	.101	II

Inspection of the data in Table 6 shows that the values for the Spring 2013 and Fall 2013 semesters, during which Format II was administered, are numerically higher than those of the first four semesters (Fall 2009 to Fall 2012), during which Format I was administered. An ANOVA test confirms that these differences are significant ($F(1, 637) = 110.4$, $p < .0005$, $\eta^2 = .148$, $\omega^2 = .146$), and that there is no violation in the assumption of homogeneity of variances ($W = 1.953$, $p_V = .163$). Hence, the format of the ABI may play a significant role on the overall reported degree of misconception endorsement.

As analyzed thus far, the large variability in the overall reported fraction of misconception endorsement may depend on the ABI format. It turns out, however, that by changing the format of the ABI, *correlations* between misconceptions remain relatively unaffected. In a paper in preparation, we will explicitly outline a method to assess these correlations. In particular, we will show that variability in the overall degree of misconception persistence has no significant influence on the correlations between misconceptions. By showing that the correlations are relatively unaffected, we can propose to group misconceptions together and sequence them in order of their *relative* difficulties, which can be used to produce orders to teach the associated concepts, from easiest to hardest, as defined by their respective mean misconception scores. These will be discussed in future papers as appropriate.

We then measured the effect of any bias from the way in which statements were phrased. To do this, we 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, we correlated the fraction of "false" statements endorsed with the fraction of "true" statements rejected (we defined what we mean by "true" and "false" on page A-25). We found that the correlation between endorsement of false statements and the rejection of true statements is .373 ($r^2 = .14$, $df = 235$, $p < .0005$). This result is statistically significant, because $p < .0005$, and is consistent with the hypothesis that students who reject correct statements are also likely to endorse misconceptions.

While the correlation above measures the strength of the tendency for students to endorse an incorrect statement or reject a true statement, an additional test is necessary to discern whether or not students endorse incorrect statements more so than they reject correct statements. A test is needed, for example, to determine if students are more likely to endorse "Earth's axis is not tilted compared to the ecliptic" than they are likely to reject "Earth's axis is tilted compared to the ecliptic." We thus use a paired-samples *t*-test (Walker, 1985, pp. 320-323) to compare the mean difference in the fraction of false statements believed vs. fraction of true statements rejected. 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 (df = 235, t = 5.77, p < .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 the next section, we provide additional support for this suggestion by testing for differences in the fractions of misconceptions endorsed between the Spring 2013 and Fall 2013 semesters.

Effect of Teaching Pedagogy

Between the Spring 2013 and Fall 2013 semesters, two different teaching pedagogies were employed, as outlined in the Method section. To examine the influence of teaching pedagogy on the endorsement of incorrect statements or rejection of scientifically-accurate statements, we first performed an ANOVA test, using data from the 129 incorrect statements, by comparing the fraction of incorrect statements believed between the Spring 2013 and Fall 2013 semesters. We present summary statistics on the fraction of incorrect statements believed in the Spring 2013 semester (n = 126, M = .302, SD = .136) and the Fall 2013 semester (n = 110, M = .264, SD = .140). 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 = .717). We found that the difference of incorrect statements believed between the Spring 2013 and Fall 2013 semesters is statistically significant ($F(1, 234) = 4.56$, p = .034, η^2 = .019, ω^2 = .015). This result is consistent with the hypothesis that *addressing misconceptions is more effective in enabling students to reduce the number of misconceptions they endorse*. The implication is that one can get a student to confront their own misinformation more effectively by teaching why the misinformation is wrong. We note, however, that the effect size is small, in the sense described in the prior section. To be clear, the small effect size does *not* discourage instruction that addresses misconceptions in *small* classroom settings. Instead, the small effect size suggests that if a researcher was to administer the ABI in two small classroom settings, each with a different instructor and pedagogy, the researcher may not obtain a statistically significant result, because the sample size may not be large enough.

For our second ANOVA test, we used a separate data set consisting of the fraction of misconceptions endorsed, associated with 86 scientifically-accurate statements, between the Spring 2013 and Fall 2013 semesters. (Because both ANOVA tests use a different data set, they are independent of each other, so a MANOVA test is unnecessary.) We present summary statistics on the fraction of scientifically-accurate statements rejected in the Spring 2013 semester (n = 126, M = .233, SD = .098) and the Fall 2013 semester (n = 110, M = .234, SD = .088). 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 = .146). We found that the difference in the fraction of endorsed misconceptions associated with scientifically-accurate statements is not at all statistically significant, i.e. $F(1, 234) = 0.001$, p = .98, η^2 = .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. The implication is that from the standpoint of teaching strictly factual information, one does not need to spend extra time teaching in the context of misconceptions, but may simply present the information as usual.

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 codes (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.

For this analysis, we chose to look at responses in the Fall 2010 semester to the following five statements (in response to discussions with colleagues who study physics education, here at the University of Maine): 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.” From the written feedback, we 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 Scantron 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 7 presents an examination of written feedback to each of the five statements, where *n* is the number of students who provided written feedback.

Table 7

Examination of Written Feedback to Five Statements of the ABI

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 7 shows that students mistakenly fill in an incorrect bubble between 0% and 9% of the time. Table 7 further shows that students misinterpret statements in the ABI 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. It is worth clarifying that these results do *not* quantify the extent to which student recollections are consistent. A detailed analysis of recollection consistency will be presented in our next paper. Of the feedback that we had available to us, according to Table 7, we 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%). Many of the incorrect “corrections” mentioned that galaxies are evenly distributed on the sky. Given that sA226 has the fourth highest degree of misconception endorsement in the ABI, our results for sA226 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.

Discussion

We performed a series of tests on the validity of the data in the ABI. From our examination, we have determined the following:

1. The presentation order of statements in the ABI has no significant influence on students’ self-reports.
2. The effect of fatigue in the process of completing the ABI has no significant influence on students’ self-reports. Hence, the interested researcher need not concern oneself with the high number of ABI statements.
3. Students taking an introductory-level course in astronomy may be more likely to endorse a misconception than they are likely to reject a scientifically-accurate statement.
4. The change in the format of the ABI due to the rephrasing of about two-fifths of the statements may cause a significant increase in the overall reported fraction of misconceptions endorsed.
5. There is a statistically significant reduction in incorrect beliefs endorsed after instruction by teaching to students in the context of their misconceptions, instead of teaching using conventional fact-oriented lecture.
6. There are no significant issues with statement misinterpretation or incorrect response codes to the associated statements. However, there may be a higher tendency for students to provide incorrect feedback to only the very hardest items in the ABI.

From the standpoint of astronomy education, the ABI presents a lot of promise, in that it can directly probe misconceptions held by students and give meaningful insights as to the persistence of misconceptions. We have determined that the data in the ABI displays convergent validity, which further suggests that there is merit in using the ABI as a tool for studying misconceptions. In our next paper, we intend to show that the tendency for one's own recollection to be inconsistent is comparable to inconsistent responses on multiple-choice tests in a longitudinal context, which suggests that the ABI is of comparable validity to multiple-choice tests.

CONCLUSION

We have introduced a new instrument consisting of a comprehensive inventory of misconceptions held by college students taking an introductory-level course in astronomy. The instrument directly probes whether or not a student believes any of the misconceptions. We find that the instrument is *not* biased in terms of the order of statement presentation or its relatively long length. It would be instructive to see how teaching in the context of misconceptions held by the students improves their grades compared to more traditional teaching.

This concludes the first paper in our series. In future papers, we will examine the consistency of student recollections of their own past beliefs; present the theoretical background for principal components analysis, a technique for identifying groups of correlated misconceptions, as the technique applies to our overall project; clarify the extent to which semester-to-semester variability in misconception endorsement influences *correlations* between misconceptions, and we will address the concern that the correlations are not significantly affected by the per-semester variability in misconception endorsement. In subsequent papers in the series, we will also construct groups of topics from the misconceptions and propose an optimal sequence to teach concepts within individual topics in astronomy.

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.

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.

Becker, L. A. (1999). *Measures of Effect Size (Strength of Association).* Retrieved May 10, 2014, http://www.uccs.edu/lbecker/glm_effectsize.html .

Bland, J. M. & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal*, *310*(6973), 170.

Borkowski, J. (2014). *Tests for homogeneity of variance*. Retrieved January 25, 2014, http://www.math.montana.edu/~jobo/st541/sec2e.pdf .

Brewin, C. R., Andrews, B., & Gotlib, I. H. (1993). Psychopathology and early experience: A reappraisal of retrospective reports. *Psychological Bulletin*, *113*(1), pp. 82-98.

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning. *American Educator*, 6-11.

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Elrbaum Associates.

Comins, N. F. (2001). *Heavenly errors: Misconceptions about the real nature of the universe*. New York: Columbia University Press.

Comins, N. F. (2014). *Heavenly errors*. Retrieved January 24, 2014, http://www.physics.umaine.edu/ncomins/ .

diSessa, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge based learning. *Cognitive Science*, *6*, 37-75.

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.

French, A., Macedo, M., Poulsen, J., Waterson, T. & Yu, A. (2008). *Multivariate Analysis of Variance (MANOVA).* Retrieved May 9, 2014, http://userwww.sfsu.edu/efc/classes/biol710/manova/MANOVAnewest.pdf .

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.

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 .

Kempton, W. (1987). Two theories of home heat control. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 222-242). Cambridge: Cambridge University Press.

Lombardi, D., Sinatra, G. M., and 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.

Olson, J. M., & Cal, A. V. (1984). Source credibility, attitudes, and the recall of past behaviors. *European Journal of Social Psychology*, *14*, 203-210.

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., and Dostal, J A. (2004). Research on a lecture-tutorial approach to teaching introductory astronomy for non-science majors. *Astronomy Education Review*, *3*, 122-136.

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.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353.

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* (pp. 3-34). New York: Routledge.

Walker, J. T. (1985). *Using statistics for psychological research: An introduction*. New York: CBS College Publishing.

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.

Andrej Favia recently completed his Ph.D. at the University of Maine, Orono, ME. He is the communicating author and can be reached at mailto:Andrej.Favia@umit.maine.edu .

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:	
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

		<u>Earth:</u>	
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
		<u>Mars:</u>	
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
		<u>Saturn:</u>	
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