Analysis of Individual *Test Of Astronomy STandards* (TOAST) Item Responses

Stephanie J. Slater, CAPER Center for Astronomy & Physics Education Research, USA
Sharon Price Schleigh, East Carolina University, USA
Debra J. Stork, University of Dubuque, USA

ABSTRACT

The development of valid and reliable strategies to efficiently determine the knowledge landscape of introductory astronomy college students is an effort of great interest to the astronomy education community. This study examines individual item response rates from a widely used conceptual understanding survey, the Test Of Astronomy Standards (TOAST). The TOAST, a 27-item, multiple-choice format, criterion-referenced test, addresses both the full range of topics commonly taught in a one- or two-semester undergraduate introductory astronomy survey courses, and concepts described in various national science education standards, frameworks, and reform documents. The present study involves an examination of responses by 1104 participants, allowing for a rigorous item-by-item and distractor-by-distractor analysis of students' responses. The results suggest that each individual TOAST item is functioning appropriately across a broad range of students, and has sufficient sensitivity to identify notable student misconceptions. These results also provide an opportunity to identify target areas of opportunity for astronomy education researchers that remain largely unstudied.

Keywords: Astronomy Education Research; Assessment; TOAST; Undergraduate Education

*M*easuring changes in college students' conceptual understanding over the duration of an introductory astronomy survey course is of widespread interest to discipline-based astronomy education researchers and teaching college professors alike. For much of the last two decades, pre- and post-course conceptual diagnostic instruments using multiple-choice response items have served as a staple of the astronomy teaching community's toolkit (Slater, S., Slater, T., Heyer, & Bailey, 2015). Wallace and Bailey (2010) also argue that well-constructed conceptual diagnostic instruments have the advantage of being able to quickly establish the range and frequency of students' ideas across astronomy. If these authors are correct in their assertions, well-designed, conceptual diagnostic efforts have great potential to provide valuable insight to teacher-educators, professional development providers, and curriculum designers who use a constructivist-oriented approach to providing and designing instruction—in other words, an approach that purposefully takes into account students' understandings and beliefs into account prior to instruction (Slater, T., Carpenter & Safko, 1996).

Identifying what the undergraduate student in an introductory astronomy course knows is important for several reasons. Students that enroll in college science introductory courses are often those that are non-science majors who will move on to become our future business leaders, politicians, journalists, historians, artists, societal leaders, parents, tax payers, voters and perhaps most importantly, teachers. It has been estimated that there are over 250,000 students who will enroll in an introductory astronomy course across the nation this year and most of these students will only take one general education course during their entire college career (Lawrenz, Huffman & Appeldoorn, 2005; Price Schleigh, 2015). Many of these students will go on to become K-12 teachers, making this work all the more critical, for these students will teach what they have learned and in the manner in which they were taught (Lawrenz, Huffman & Appeldoorn, 2005; Price Schleigh et al. 2011).

The *Test Of Astronomy STandards*, known more commonly as the TOAST, is perhaps the most recently developed instrument in astronomy education research enjoying widespread adoption. Described in detail elsewhere

by S. Slater (2014), the TOAST was designed and validated by using recommended principles (Slater, S., Slater, T., & Shaner, 2008; Slater, S., 2009). As a 27-item, multiple-choice format assessment instrument, the TOAST addresses both the full range of topics commonly taught in a one- or two-semester undergraduate introductory astronomy survey course, and concepts described in national science education standards, frameworks, and reform documents. In contrast to many contemporary assessment efforts, the TOAST is a criterion-referenced test (CRT) rather than norm-referenced test (NRT). CRTs are primarily used to characterize student learning, as compared to externally described learning objectives while NRTs are designed to intentionally highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement, from high achievers to low achievers (Stiggins, 1994). NRTs use a representative group of students as a baseline prior to availability to the public, and are used to divide students into groups based upon student performance. The scores of the students who take the test after publication are then compared to those of the norm-baseline group, usually for a period of several years. As such, students who answer the majority of items on an NRT correctly may still be ranked poorly if all other students also performed well. As a CRT, the TOAST does not compare and rank students, but is best used to compare individual performance to those astronomical learning objectives previously defined by astronomy community consensus (Slater, S., 2014). In an ideal world, it is hope that all students can and will perform well on the TOAST.

In his award winning paper, Sadler (1998) proposed that the most insightful multiple-choice items used for research on student understanding were those that were psychometrically driven. In this sense, he meant that psychometrically driven items were ones that tapped previously established misconceptions widely held by students and suggested that the multiple-choice distractors offered to students be those that have been established by systematic astronomy education research (Sadler et al., 2010). This philosophical perspective has influenced the development of a generation of assessment instruments in astronomy education and astronomy education research, most particularly the TOAST (Slater, S., 2014). In further response to Sadler's proposal that the most useful measurement items are based upon pre-existing research, this paper provides a detailed item-by-item and distractor-by-distractor analysis of students' responses, comparing students' responses on the TOAST to the extant education research literature on students' astronomical misconceptions.

METHOD

Participants

In order to determine the extent to which the individual TOAST items were sufficiently sensitive to successfully diagnose commonly cited astronomy misconceptions from the literature, responses were studied from 1,104 undergraduate students taking introductory astronomy survey courses designed for non-science majors at five different institutions, spread across the United States. These students took the survey voluntarily and confidentially at the beginning of the class before the instructors had taught any astronomy content, as dictated by the approved IRB human subjects plan. It is assumed that most of these students were exposed to the astronomy concepts listed in national standards and frameworks documents during their K-12 experience, but there is no way to measure this. Responses were visually inspected and student responses with more than two missing pieces of information were removed from the sample. The remaining sample included 1,066 responses.

TOAST Item Sensitivity

Varma (2008) recommends conducting a detailed item-by-item analysis of multiple-choice items created for surveying conceptual understanding. An abridged analysis of TOAST items was previously reported by S. Slater (2014), but because responses for different populations may vary, data from these participants were analyzed using *Remark Classic OMR v2* to calculate Cronbach alpha values, inspect item difficulty levels and item discrimination indices. The Cronbach alpha score is a measure of coherence, which is widely judged to indicate the presence of internal reliability and a lack of user test fatigue (Nunnally, 1978; Vasser & Crosby, 2008). The Cronbach alpha observed for these college students of $\alpha = 0.83$ is interpreted to mean that 83% of the variation in the results represents true variance rather than error variance. The widely accepted Cronbach alpha cut-off is 0.70 or higher for a set of items to be considered to demonstrate a sufficiently high level of internal consistency.

To further establish sufficient sensitivity of the instrument, two other aspects of Classic Test Theory were also analyzed by: item difficulty and item discrimination. Item difficulty is a measure of the proportion of the population who correctly answered the test question. As such, this statistic might better be known as item "easiness" rather than "difficulty" because a high value means that most respondents got the answer correct. This data is shown in Table 1. Haladyna, Downing, and Rodriguez (2002) argue that an item difficulty percentage between .30 and .90 is most desirable. The average item difficulty (p-value) on TOAST items for this sample is 0.46 with all items scoring in a desirable range. P-values for each TOAST item, for this sample, are given in Table 1.

Item discrimination is most often defined as a measure to which success on a given item equates to a respondents' success score on the overall instrument. A high item discrimination value means that respondents who do well overall, tend to answer the specific item correctly as well. If students who do well overall tend to answer an item incorrectly, the item will have a zero or even a negative item discrimination index. An item discrimination of 0.15 or higher is most often considered satisfactory (Nunnally, 1978). The average item discrimination index on the TOAST overall was 0.42 with 0.28 representing the lowest value for any one item. Item discrimination for each item is given in Table 1 as a correlation, calculated as the Pearson correlation between responses to a particular item and scores on the total test.

Collectively, these indicators suggest that each item is functioning as intended and makes a meaningful contribution to the overall TOAST score.

Table 1. Item Difficulty and Discrimination for TOAST Test Items

|	Item Difficulty	Item Discrimination		Item Difficulty	Item Discrimination
Item 1	0.37	0.44	Item 15	0.28	0.41
Item 2	0.39	0.43	Item 16	0.79	0.42
Item 3	0.57	0.39	Item 17	0.28	0.29
Item 4	0.66	0.46	Item 18	0.56	040
Item 5	0.61	0.33	Item 19	0.22	0.56
Item 6	0.23	0.38	Item 20	0.35	0.42
Item 7	0.53	0.28	Item 21	0.33	0.47
Item 8	0.43	0.58	Item 22	0.40	0.44
Item 9	0.47	0.32	Item 23	0.40	0.50
Item 10	0.63	0.40	Item 24	0.28	0.41
Item 11	0.60	0.39	Item 25	0.41	0.48
Item 12	0.36	0.50	Item 26	0.20	0.31
Item 13	0.32	0.61	Item 27	0.26	0.37
Item 14	0.39	0.43			

TOAST Construct Validity

As a means of judging this instrument's construct validity, the distribution of participants' responses to each item were compared to the literature on student learning on that topic. In order for the TOAST to function in a theoretically valid way, it is expected that student responses to items addressing topics in which there are significant and robust literature on student misconceptions should show a distribution reflecting those misconceptions. In some cases, there is a robust extant literature on the ways in which many student populations engage with the content (e.g., seasons, lunar phases). Findings from this literature base were compared to TOAST item results. In other cases, there is virtually no peer-reviewed empirical findings related to the underlying mechanisms and beliefs that students bring to the astronomy content (e.g., light, spectra, stars, cosmology). While there is literature describing the kinds of misconceptions that students possess in these areas, there is little or no research empirically substantiating students' mental models. In these instances, the item results are compared to other areas of research (e.g., visualizations) and to common observations of teaching faculty. These gaps were interpreted to indicate potential, fruitful areas for further research.

RESULTS

The mean pre-course score for the analyzed college student population in the United States sampled was 12 out of 27 possible, or 44%, while the median was 11 points, and the standard deviation was 5.4 (n=1,066). A detailed report of the frequency distribution and item discrimination values for each TOAST item is given below.

Item 1

Figure 1. Object for Items 1 and 2
Use the drawing below to answer the next two questions.
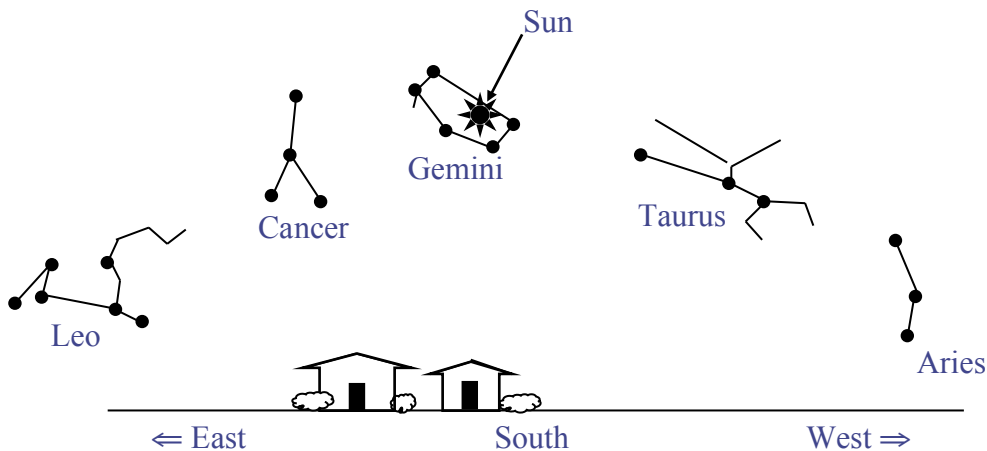
Figure 2. Item 1 with Analysis

Using the drawing above: If you could see stars during the day, the drawing above shows what the sky would look like at noon on a given day. The Sun is at the highest point that it will reach on this day and is near the stars of the constellation Gemini. What is the name of the constellation that will be closest to the Sun at sunset on this day?

Response Labels	Percent college students (253)	Item Discrimination Index
a. Leo	11.07	-0.14
b. Taurus	2.77	-0.11
c. Aries	38.34	-0.22
d. Cancer	2.77	-0.11
e. Gemini	*36.76*	*0.44*

Figure 3. Item 2 with Analysis

Using the drawing above: This picture shows the position of the stars at noon on a certain day. How long would you have to wait to see Gemini at this same position at midnight?

| Response Labels	Percent college students (253)	Item Discrimination Index
a. 12 hours	17.79	-0.15
b. 24 hours	7.51	-0.16
c. 6 months	*39.13*	*0.43*
d. 1 year	6.72	-0.11
e. Gemini is never seen at this position at midnight.	18.18	-0.18

Vosniadou and Brewer (1992) reported that K-12 students often believe that the stars in the sky (other than the sun) are "fixed and unmoving" in the sky. This work was corroborated by Plummer (2009) who found that the belief that the stars do not move across the sky is pervasive, with 65% of her sample of U.S. eighth graders describing a fixed-star sky. In the results for TOAST Item 1 (based on the illustration shown in Figure 1 and data shown in Figure 2), we see that although there are nearly the same number of responses between Distractor C and

the scientifically acceptable Answer E, with this sample of U.S. college students continuing to prefer the "fixed" notion of the stars. This result is consistent with students' immature understanding of the observable consequences of Earth's rotation reported across the literature and summarized by S. Slater (Slater Parker, 2006). Distractors B and D do not serve to provide insight into student misconceptions, but are instead present to provide item face validity to survey participants. It is unclear why students prefer Distractor A so much more than B or D. This appears to be a rich area for further investigation.

In the results for Item 2, participants preferred Answer C, although a nearly equal number of respondents split their preferences between Distractors A and E. Distractor A appeals to respondents who, as in the previous item, hold a "fixed" star mental model of the sky, in which the stars do not appear to move across the sky in the same way the Sun does, providing corroborating data for that seen in Item 1. Participants preferred Distractor E to Distractors B and D, which serve to provide item face validity to survey participants; the reason for this preference is unclear.

Item 3

Figure 4. Object for Item 3
Use the drawing below to answer the next question.
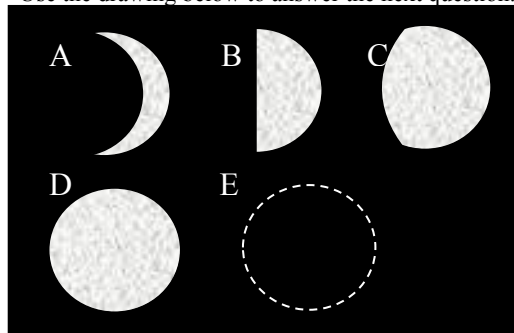
Figure 5. Item 3 with Analysis

You look to the eastern horizon as the Moon first rises and discover that it is in the new moon phase. Which picture shows what the moon will look like when it is at its high point in the sky, later that same day?

Response Labels	Percent college students (253)	Item Discrimination Index
a. A	12.65	-0.16
b. B	6.32	-0.14
c. C	4.74	-0.19
d. D	12.65	-0.17
e. E	*57.31*	*0.39*

In the results for Item 3, respondents preferred Answer E, although a number of respondents who are attracted to Distractors A and D. Both choices are believed to appeal to respondents who have a mental model that the moon's appearance changes significantly during each Earth rotation (Lindell, 2004; Plummer, 2009). It is unclear why participants prefer Distractors A and D to Distractors B and C. The research on moon phases described elsewhere by Lindell (2004) did not deeply probe this particular belief in fine detail, leaving substantial room for future work to be done.

Item 4

Figure 6. Item 4 with Analysis

You are located in the continental U.S. on the first day of October. How will the position of the Sun at noon be different two weeks later?

Response Labels	Percent college students (253)	Item Discrimination Index
a. It will have moved toward the North.	3.95	-0.16
b. It will have moved to a position higher in the sky.	12.25	-0.17
c. It will stay in the same position.	9.49	-0.29
d. It will have moved to a position closer to the horizon.	*65.61*	*0.46*
e. It will have moved toward the west.	3.95	-0.13

In the results for Item 4 shown in Figure 6, participants preferred Answer D, with a smaller group of respondents splitting their preferences between Distractors B and C. (It should be noted that Distractor A is essentially the same answer as Distractor B, and any analysis that tries to make meaning of results from this item should take that into account.) These distractors indicate sampled respondents often have an unclear understanding of the changes in the observable sky through the seasons, and that a noteworthy percentage of college students continue to believe that the Sun is higher in sky during the Northern Hemisphere winter. This agrees with Plummer's finding that many students entering high school describe a winter sun that is higher in the sky than a summer sun, despite living in a middle latitude location where the difference in the Sun's position is quite noticeable through the course of the season (2009). This result agrees with findings such as those reported decades earlier by Sadler (1992) and S. Slater (2006).

Item 5

Figure 7. Item 5 with Analysis

Which sentence best describes why the Moon goes through phases?

Response Labels	Percent college students (253)	Item Discrimination Index
a. Earth's shadow falls on different parts of the Moon at different times.	13.04	-0.18
b. The Moon is somewhat flattened and disk-like. It appears more or less round depending on the precise angle from which we see it.	11.46	0
c. Earth's clouds cover portions of the Moon resulting in the changing phases that we see.	5.93	-0.15
d. The sunlight reflected from Earth lights up the Moon. It is less effective when the Moon is lower in the sky than when it is higher in the sky.	6.72	-0.21
e. We see only part of the lit-up face of the Moon depending on its position relative to Earth and the Sun.	*60.87*	*0.33*

In the results for Item 5 shown in Figure 7, participants preferred Answer E, with a smaller group of respondents showing an attraction to Distractor A. This distractor attracts respondents as predicted in the literature. This corroborates earlier studies by Baxter (1989), Skam (1994) and Dai (1991), among others, who all indicate that many individuals believe that lunar phases are caused by the Earth's shadow falling on the moon. The selection of all distractors provides an interesting comparison to the data collected by Bisard et.al. (1994) in which the authors found that the scientifically correct response was narrowly preferred to a choice that described phases as being caused by Earth's shadow, which was chosen by 37.6% of respondents. In that study 18.8% of the students chose the distractor indicating that the cause of the lunar phases was the varying angle of sunlight off the Earth, while the smallest percentage of students, 4.4%, chose clouds as being the cause of the lunar phases.

Item 6

Figure 8. Item 6 with Analysis

Imagine you see Mars rising in the east at 6:30 pm. Six hours later what direction would you face (look) to see Mars when it is highest in the sky?

Response Labels	Percent college students (253)	Item Discrimination Index
a. toward the north	17	-0.17
b. toward the south	*23.32*	*0.38*
c. toward the east	2.77	-0.17
d. toward the west	5.93	-0.12
e. directly overhead	39.53	-0.04

In the results for Item 6, participants showed a very small preference for Answer B as measured by the response receiving the greatest number of choices, with a larger group of the respondents choosing Distractor E. The results for Item 6 corroborate the results from Item 4. In both cases respondents are asked to predict the transit of an object at a certain time, and in both cases respondents made incorrect predictions, with some respondents predicting that Mars' transit would occur to the north, and a larger group predicting transit at the zenith. These predictions are only true for a very narrow region on Earth and few participants are believed to originate from this region, while all of the participants were outside of this region at the time of testing. Therefore, this result is more likely a result of respondents' poor knowledge of the motions of objects in the celestial sphere, due to a lack of observation and an incomplete understanding of the consequences of Earth's rotation on a tilted axis.

Item 7

Figure 9. Item 7 with Analysis

Imagine that Earth was upright with no tilt. How would this affect the seasons?

Response Labels	Percent college students (253)	Item Discrimination Index
a. We would no longer experience a difference between the seasons.	*52.96*	*0.28*
b. We would still experience seasons, but the difference would be *less* noticeable.	22.92	-0.05
c. We would still experience seasons, but the difference would be *more* noticeable.	13.04	-0.21
d. We would continue to experience seasons in essentially the same way we do now.	4.35	-0.11

In the results for Item 7, respondents' preferred Answer A, with a smaller group of respondents splitting their preferences between Distractors B and C. These distractors indicate most respondents hold an unclear understanding of the cause of Earth's seasons. Sadler (1992) and S. Slater (Parker, 2006) reported that even after direct instruction in K-12 classrooms that specifically address the relationship between seasons and Earth's tilt, students continue to hold onto these misconceptions (viz., Atwood & Atwood, 1996; Schneps, 1988; and Kikas, 2004 for earlier research on this topic). These results suggest that many undergraduate science students, despite their K-12 educations, continue to have an unclear understanding about the role that Earth's tilt plays in seasonal changes.

Item 8

Figure 10. Item 8 with Analysis

How does the Sun produce the energy that heats our planet?

Response Labels	Percent college students (253)	Item Discrimination Index
a. The gases inside the Sun are burning and producing large amounts of energy.	18.97	-0.26
b. Gas inside the Sun heats up when compressed, giving off large amounts of energy.	11.46	-0.15
c. Heat trapped by magnetic fields in the Sun is released as energy.	10.67	-0.18
d. Hydrogen is combined into helium, giving off large amounts of energy.	*43.08*	*0.58*
e. The core of the Sun has radioactive atoms that give off energy as they decay.	8.3	-0.14

In the results for Item 8, participants sampled largely preferred Answer D, with smaller groups of respondents splitting their preferences between Distractors A, B and C. Distractor A speaks to a widespread belief that the sun produces light through the burning of material. Distractors B and C speak to other known methods of producing heat, which are probably reflective of everyday experience. These alternative beliefs have been extensively documented by Agan (2004) and Bailey (2007) and are generally consistent with the results of this study.

Item 9

Figure 11. Item 9 with Analysis

The Big Bang is best described as:

Response Labels	Percent college students (253)	Item Discrimination Index
a. The event that formed all matter and space from an infinitely small dot of energy.	*46.64*	*0.32*
b. The event that formed all matter and scattered it into space.	10.67	-0.22
c. The event that scattered all matter and energy throughout space.	23.32	-0.04
d. The event that organized the current arrangement of planetary systems. to further establish the sufficient sensitivity of the instrument	10.67	-0.15

In the results for Item 9, respondents largely preferred Answer A, with other respondents splitting their preferences between Distractors B, C and D. This result indicates that each distractor is relatively attractive to respondents. This item is one of four that was created specifically for the TOAST rather than being adapted from other instruments. This item is based upon work that investigated the nature of students' beliefs related to the Big Bang (Hansson & Redfors, 2006; Prather, Slater, T., & Offerdahl, 2002; Wallace, 2011) which investigated the nature of students' beliefs. These papers reported the dominance of the thought that the Big Bang was an event involving pre-existing empty space, with many students also believing that the Big Bang involved the arrangement of existing matter, including large scale objects (e.g., planets). Based upon the reported findings, and the results of this item, there is reason to believe that students' conceptions of the Big Bang is an area that deserves additional attention in future studies.

Item 10
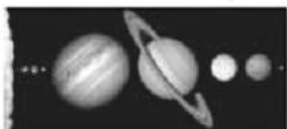
Figure 12. Item 10 with Analysis

Which of the following ranks locations, from closest to Earth to farthest from Earth?

Response Labels	Percent college students (253)	Item Discrimination Index
a. the Sun, the Moon, the edge of our solar system, the North Star, the edge of our galaxy	4.35	-0.13
b. the Sun, the North Star, the Moon, the edge of our galaxy, the edge of our solar system	2.37	-0.19
c. the Moon, the North Star, the Sun, the edge of our solar system, the edge of our galaxy	18.58	-0.21
d. the Moon, the Sun, the edge of our solar system, the North Star, the edge of our galaxy	*63.24*	*0.40*
e. the North Star, the Moon, the Sun, the edge of our galaxy, the edge of our solar system	5.53	-0.14

In the results for Item 10, respondents preferred Answer D, with a smaller group of students preferring Distractor C. Overall, the scores on this item strongly suggest that many participants believe the North Star, and by extension other stars, are closer to the Earth than the Sun. This is consistent with work reported by S. Slater, Morrow, and T. Slater (2008) that K-12 students struggle to understand basic astronomical geography. Fanetti (2001) found similar results in the college population. Her work strongly supports the notion that college students' inaccurate mental models of the Sun-Earth, Moon system—and particularly their relative sizes and distances—prevent them from being able to accurately reasoning about this concept. This could also be interpreted as being consistent with findings that both K-12 teachers and K-12 students conceive of the Sun and stars as fundamentally different objects. (Slater, T. 1993; Turkoglu, Ornek, Gokdere, Suleymanoglu, & Orbay, 2009; Vosniadou & Brewer, 1992). In Vosniadou's and Brewer's work (1992), they found that the students she studied articulate that the Sun is a star, as a piece of cultural knowledge, but that upon further questioning, describe the Moon as being more star-like than the Sun. Without a proper understanding of the nature of these objects, such individuals are not able to appropriately apply ideas related to size, distance, and appearance.
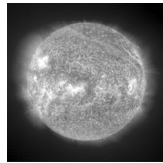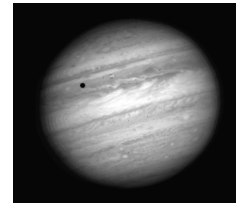
Item 11

Figure 13. Object for Item 11
Consider the six different astronomical objects (A-F) shown below.
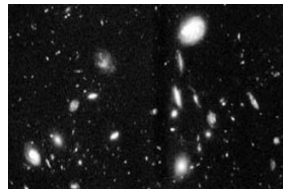

A. The Solar System

B. The Sun

C. Jupiter

D. Andromeda Galaxy

E. Galaxy Cluster

F. Nebula

Figure 14. Item 11 with Analysis

Which of the following is the best ranking (from smallest to largest) for the size of these objects?

Response Labels	Percent college students (253)	Item Discrimination Index
a. C<F<B<A<D<E	9.09	-0.15
b. E<D<F<A<B<C	10.28	-0.19
c. C<B<A<F<D<E	***60.08***	***0.39***
d. F<C<B<A<D<E	8.3	-0.25
e. None of the above is correct.	7.11	-0.01

In the results for Item 11, nearly 60% of respondents could correctly rank astronomical objects by size. The results for the remaining 40% of respondents do not clearly end up in well-understood categories, suggesting respondents might be simply guessing on this. At this time there is some research that points to a basic lack of student understanding of astronomical geography as described elsewhere by Slater and Morrow (2010), but astronomical geography and scale are clearly areas deserving further research.

Item 12

Figure 15. Item 12 with Analysis

Imagine that Earth's orbit were changed to be a perfect circle about the Sun so that the distance to the Sun never changed. How would this affect the seasons?

Response Labels	Percent college students (253)	Item Discrimination Index
a. We would not be able to notice a difference between seasons.	18.7	-0.31
b. The difference in the seasons would be *less* noticeable than it is now.	27.83	-0.09
c. The difference in the seasons would be *more* noticeable than it is now.	13.04	-0.22
d. We would experience seasons in the same way we do now.	***40.43***	***0.5***

In the results for Item 12, respondents' preferred Answer D, with a larger group of respondents splitting their preferences between Distractors A, B and C. This result is similar to the result for Item 7 and serves to corroborate that interpretation.

Item 13

Figure 16. Item 13 with Analysis

What is a star?

Response Labels	Percent college students (253)	Item Discrimination Index
a. a ball of gas that reflects light from another energy source	7.51	-0.14
b. a bright point of light visible in Earth's atmosphere	2.77	-0.08
c. a hot ball of gas that produces energy by burning gases	35.18	-0.36
d. a hot ball of gas that produces energy by combining atoms into heavier atoms	***32.02***	***0.61***
e. a hot ball of gas that produces energy by breaking apart atoms into lighter atoms	11.86	-0.08

In the results for Item 13, respondents' preferred Distractor C to the scientifically accurate Answer D. The item discrimination indices suggest that the item is sufficiently sensitive for measuring understanding and is functioning in an appropriate manner; respondents who seem to have a good overall understanding of astronomy content, as measured by their overall score, tend to choose the correct answer. The negative discrimination score for Distractor C indicates that many students who perform well overall choose this incorrect response. As in the results for Item 8, Distractor C speaks to the historical belief that the sun produces light through the burning of material reported earlier by Bailey (2007). Distractors B and E speak to other known methods of producing heat, in common

experience, which is in line with Agan's findings that many students believe that the Sun is made of fire or "lava" (2004).

Item 14

Figure 17. Item14 with Analysis

Which one property of a star will determine the rest of the characteristics of that star's life?

Response Labels	Percent college students (253)	Item Discrimination Index
a. brightness	12.65	-0.23
b. temperature	19.37	-0.19
c. color	4.35	-0.03
d. mass	**39.13**	**0.43**
e. chemical makeup	13.04	-0.13

   Item 14 is sourced from the Star Properties Concept Inventory (Bailey, 2007). For Item 14, the scores suggest that participants are largely guessing rather than revealing any detailed understanding or mental models they hold. This is consistent with earlier work, suggesting that students have few strongly held conceptual beliefs in this domain. This corroborates work by Comins (2001) who examined the misconception that the sun/stars will last forever and found that students easily changed their beliefs. Agan (2004) also suggested that undergraduates' misconceptions for stellar evolution and stellar cycles needs to be addressed early on during instruction as this thinking interferes with their ability to develop mental models that involve relationships between stellar characteristics and stellar life cycles. While these results indicate that item is functionally adequately, Bailey's (2007) thorough description of students' responses to questions related to stars does not speak to the question of students' mental mechanisms, indicating that this may be a fruitful area for further research.

Item 15

Figure 18. Item 15 with Analysis

Current evidence about how the universe is changing tells us that

Response Labels	Percent college students (253)	Item Discrimination Index
a. We are near the center of the universe.	17.86	-0.21
b. Galaxies are expanding into empty space.	30.36	-0.12
c. Groups of galaxies appear to move away from each other	**32.81**	**0.41**
d. Nearby galaxies are younger than distant galaxies.	14.73	-0.14

   Item 15 is one of four that was created specifically for the TOAST and little comparative information exists in the science education literature. This item is designed to probe students' conceptions of the cosmological expansion of the universe. The work started by Prather, Slater, and Offerdahl (2002) and the much more recent work of Wallace (2011) do not delve deeply into this aspect of cosmology. Therefore, item development relied upon the collective expertise of longtime astronomy instructors to predict ways in which respondents' phenomenological primitives or misconceptions and their acquired cultural knowledge might interact to construct erroneous synthetic notions. Distractor A is intended to attract respondents who have translated the statement that all galaxies appear to be moving away from ours into a notion that we must be at the center of the universe. Distractor B is intended to attract respondents who believe that all galaxies are expanding outward from some central point, into pre-existing, empty space. Data related to Distractor B is reminiscent of data collected in Item 9, in which a similar percentage of respondents indicated a notion that includes matter moving into pre-existing space. Distractor D is intended to attract respondents who have heard of the concept of "look-back time" but have not integrated the idea enough to be able to correctly apply it to the problem of expansion. In contrast, Distractor C does not support the notion of pre-existing, empty space, but does reflect a correct understanding of gravitationally bound units. This is a conceptual domain is still fruitful for future astronomy education researchers.

Item 16

Figure 19. Item 16 with Analysis

Stars begin life as

Response Labels	Percent college students (253)	Item Discrimination Index
a. a piece off of a star or planet.	3.16	-0.07
b. a white dwarf.	7.11	-0.28
c. matter in Earth's atmosphere.	5.53	-0.16
d. a black hole.	1.98	-0.16
e. a cloud of gas and dust.	*79.45*	*0.42*

Item 16 appears to divide students into two groups: those who knew the scientifically correct answer and those who did not. This question is sourced from the Star Properties Concept Inventory (SPCI) (Bailey, 2007). In this case, it appears that the item is testing for knowledge level recall rather than a deep conceptual idea. Bailey's doctoral work did not probe for students' underlying mental mechanisms related to this item, indicating an area for further investigation.

Item 17

Figure 20. Item 17 with Analysis

When the Sun reaches the end of its life, what will happen to it?

Response Labels	Percent college students (253)	Item Discrimination Index
a. It will turn into a black hole	25.69	-0.21
b. It will explode, destroying Earth	26.09	-0.02
c. It will lose its outer layers, leaving its core behind	*28.46*	*0.29*
d. It will not die due to its mass	8.97	-0.13

The results for Item 17 are strikingly similar to the other TOAST items related to the work of Bailey (2007) on the Sun, suggesting that this item is performing as expected. The results here are interpreted to mean that targeted instruction on the processes governing the Sun is largely an underserved area of teaching at the K-12 level.

Item 18

Figure 21. Item 18 with Analysis

If you were in a spacecraft near the Sun and began traveling to Pluto you might pass

Response Labels	Percent college students (253)	Item Discrimination Index
a. planets.	1.58	-0.07
b. stars.	2.77	-0.08
c. moons.	2.77	-0.13
d. two of these objects.	*56.13*	*0.40*
e. all of these objects.	30.83	-0.31

Item 18 is one of four, created specifically for the TOAST. It is designed to probe participants' conceptions of the makeup of the solar system. Respondents who selected Distractor E indicated that they conceive of a solar system with containing planets, moons, and stars. Previous research by S. Slater, Morrow, and T. Slater (2008) indicates that many high school students believe that there are many stars in our solar system and that other stars, such as the North Star, are closer to Earth than the Sun. Similarly, Agan (2004) found that students often conceive of stars are small nearby objects, rather than large, distant objects. These results suggest that this item may be to be able to successfully discriminate between participants who know that there is only one star in our solar system, and those who think that there are many. Unfortunately, it is unclear whether respondents who chose Answer D are indicating that they know that there are no other stars in the solar system, or that they believe the solar system to lack either planets or moons. This uncertainty provides cause for further research.

Item 19

Figure 22. Item 19 with Analysis

How did the system of planets orbiting the Sun form?		
Response Labels	**Percent college students (253)**	**Item Discrimination Index**
a. The planets formed from the same materials as the Sun.	*22.92*	*0.56*
b. The planets and the Sun formed at the time of the Big Bang.	18.18	-0.18
c. The planets were captured by the Sun's gravity.	33.6	-0.22
d. The planets formed from the fusion of hydrogen in their cores.	11.86	-0.16

   Item 19 is one of four that was created specifically for the TOAST. It is designed to probe respondents' conceptions of the formation of the solar system. Distractor B was designed to attract respondents who express a conception of the Big Bang in which solid materials were explosively ejected during the Big Bang (Prather, Slater, & Offerdahl, 2002). This conception was also elicited in Item 9. Distractor C represents an older, failed hypothesis of the formation of our planetary system. Distractor D was designed to appeal to respondents who have fractured cultural knowledge, confounding scientific ideas related to "cores" and "fusion" (Vosniadou & Brewer, 1992).

Item 20

Figure 23. Item 20 with Analysis

Which of the following would make you weigh half as much as you do right now?		
Response Labels	**Percent college students (253)**	**Item Discrimination Index**
a. Take away half of the Earth's atmosphere.	21.46	-0.13
b. Double the distance between the Sun and the Earth.	18.91	-0.14
c. Make the Earth spin half as fast.	14.45	-0.07
d. Take away half of the Earth's mass.	*45.18*	*0.42*

   Results from Item 20 indicate that participants largely are able to identify the scientifically accurate Answer D to the offered distractors. Distractors A, B, and C are all based on findings reported by Treagust and Smith (1989), Osborne and Gilbert (1980), Philips (1991), Clark and colleagues (2014), and Schleigh and colleagues (2015) which indicate that many K-12 students believe that gravitational attraction or force is related to the presence of air; motion or the speed of a planet's rotation; or the distance between the planet and the Sun. These results indicate that college students are attracted to these ideas at fairly similar rates. This item was modified from an earlier item used on the Astronomy Diagnostics Test 2 (ADT2) (Zeilik, 2002). For the TOAST, the item was modified to remove Distractor E, which initially read: "More than one of the above." This older distractor was removed, as the research team had no way to discern which of the responses students might be selecting. This modified version of the item for the TOAST forces respondents to select the response that most closely aligns with mental models held by the general population.

Item 21

Figure 24. Item 21 with Analysis

Astronauts "float" around in the space shuttle as it orbits Earth because		
Response Labels	**Percent college students (253)**	**Item Discrimination Index**
a. There is no gravity in space	28.92	-0.13
b. They are falling in the same way as the Space Shuttle	*40.16*	*0.47*
c. They are above earth's atmosphere	19.34	-0.07
d. There is less gravity inside of the Space Shuttle	11.58	-0.15

   Results from Item 21 indicate that more than half of participants sampled prefer the scientifically accurate Answer B to the offered distractors. Distractors A, C, and D are based on the same literature base that was cited in the analysis of Item 20. Note that, like Item 20, this item was modified from an item currently available on the ADT2; Distractor E, "more than one of the above," was removed.

Item 22

Figure 25. Item 22 with Analysis

Energy is released from atoms in the form of light when electrons

Response Labels	Percent college students (253)	Item Discrimination Index
a. are emitted by the atom.	17	-0.21
b. move from low energy levels to high energy levels.	25.3	-0.15
c. move from high energy levels to low energy levels.	*45.06*	*0.44*
d. move in their orbit around the nucleus.	7.11	-0.15

   For this item, participants performed well, with nearly one half of respondents answering correctly. This TOAST item was based on work reported earlier by Bardar (Weeks) and colleagues (2006). However, Bardar's work does not specifically cite a reason why respondents might be selecting the distractors that they do. This provides a potentially fruitful area for further research.

Item 23

Figure 26. Item 23 with Analysis

Which of the following would be true about comparing visible light and radio waves?

Response Labels	Percent college students (253)	Item Discrimination Index
a. The radio waves would have a lower energy and would travel slower than visible light.	12.65	-0.04
b. The visible light would have a shorter wavelength and a lower energy than radio waves.	11.46	-0.2
c. The radio waves would have a longer wavelength and travel the same speed as visible light.	*40.32*	*0.50*
d. The visible light would have a higher energy and would travel faster than radio waves.	19.37	-0.27
e. The radio waves would have a shorter wavelength and higher energy than visible light.	7.91	-0.09

   The responses to Item 23 reveal that all participants struggle with a scientifically accurate mental model for the nature of light. In ADT2 validation results, T. Slater and colleagues (1999) reported that many respondents readily confuse the nature of light waves with the nature of acoustic sound waves, even though they are fundamentally different.

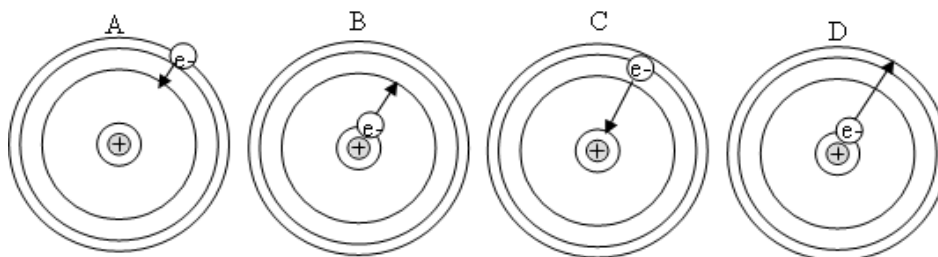**Item 24**

Figure 27. Item 24 with Analysis

The atoms in the plastic of your chair were formed

Response Labels	Percent college students (253)	Item Discrimination Index
a. in our Sun.	9.88	0.04
b. by a star existing prior to the formation of our Sun.	*27.67*	*0.41*
c. at the instant of the Big Bang.	32.81	-0.18
d. approximately 100 million years ago.	12.65	-0.22
e. in a distant galaxy in a different part of the early universe.	5.53	-0.05

   Item 27 is the fourth item specifically designed for the TOAST, and is intended to probe student thinking related to solar system and planetary formation. The results for Item 24, distractor C, which relates to the formation of heavy elements in the instant of the Big Bang, was a more popular choice. The idea that heavy elements, solid materials, and even fully formed objects are related to the Big Bang event was elicited in Items 9 and 19 at a frequency similar to the result seen here. Another interesting detail in this data is related to Distractor A. Distractor A was designed to attract respondents who may misapply the culturally transmitted idea in which heavy element formation occurs, in certain situations, in the cores of stars. This option was not chosen at a high frequency, but the respondents who did choose this answer performed moderately well on the remainder of the instrument. This makes

sense; one must possess cultural knowledge before they are able to apply it in inappropriate ways. The results on this item indicate that it is a fruitful area for future astronomy education research.

Items 25 and 26

Figure 28. Objects for Items 25 and 26
Use the drawings below to answer the next two questions.

Figure 29. Item 25 with Analysis

Which atom would be absorbing light with the greatest energy?

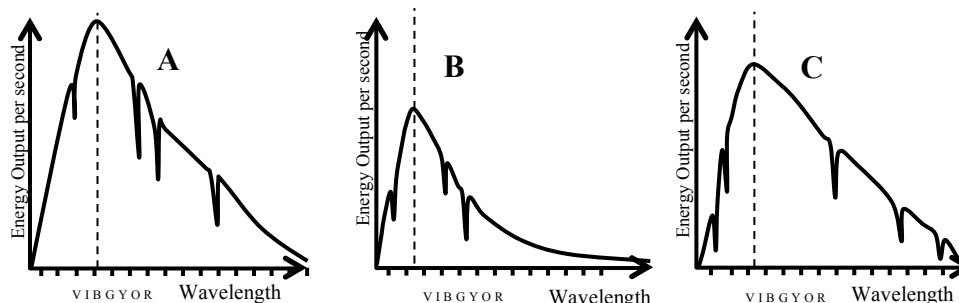| Response Labels	Percent college students (253)	Item Discrimination Index
a. A	10.67	-0.26
b. B	12.65	-0.26
c. C	27.27	-0.08
d. D	*41.11*	*0.48*

Figure 30. Item 26 with Analysis

Which atom would emit light with the shortest wavelength?

Response Labels	Percent college students (253)	Item Discrimination Index
a. A	26.09	0.14
b. B	28.06	-0.22
c. C	*19.76*	*0.31*
d. D	16.21	-0.13

In Items 25 and 26, respondents are required to correctly apply their understanding of light production to the actions occurring within four generic atoms. The accuracy of respondents' responses should not depend on whether respondents' conceive of the circles as energy levels or electron orbitals. This item was sourced from the LSCI, and while results from Barder, Prather, Brecher, and T. Slater (2006) and this study, suggest that the items are functioning appropriately, the source material does not indicate rationales for the distractors. A best guess might suggest that participants are misapplying a range of iconic graphical interpretations.

Item 27

Figure 31. Objects for Item 27
Use the graphs below to answer the next question.

Figure 32. Item 27 with Analysis

The graphs below (above) illustrate the energy output versus wavelength for three unknown objects A, B, and C. Which of the objects has the highest temperature?

Response Labels	Percent college students (253)	Item Discrimination Index
a. A	35.18	-0.14
b. B	*25.69*	*0.37*
c. C	8.3	-0.18
d. All three objects have the same temperature.	6.72	-0.05
e. The answer cannot be determined from this information.	11.46	-0.08

Results for Item 27 (Figures 31 & 32) indicate that this is a challenging item for participants. This item is also sourced from the LSCI. As in the cases of previous LSCI items, this item appears to function appropriately, but source documents do not provide an underlying rationale for the distractors. Respondents prefer Distractor A to the scientifically accurate Answer B. It is possible that respondents are incorrectly reading the graph, confusing "energy output per second" with flux, and then erroneously applying the Stefan-Boltzmann relationship to determine that Star A must be the star with the highest temperature. This interpretation would require that those same respondents ignore the peak wavelength of each star (and Wien's Law). In addition, the item discrimination index calculation indicates that there is an inverse relationship between high overall scores on the instrument and the choice of Distractor A. In other words, the respondents who earned the highest scores on this test of general astronomy knowledge did not choose Distractor A.

We speculate that students are reasoning through the questioning using a variety of misapplied iconic representational schemes. For instance, as the question asks for the star with the highest temperature, students respond by selecting the graph with the "highest" curve. Further research into students' understanding of this topic may benefit by considering students' general graphical reasoning.

DISCUSSION

TOAST is unique in that as a criterion referenced test (CRT), it views astronomy content knowledge through a standards-based lens. The knowledge that the participants demonstrate is not random but directly relates to what experts in the field of astronomy education have come to consensus that all K-12 students, therefore all American citizens should possess. It also confirms the research that suggests that students in the K-12 classrooms continue to hold onto their misconceptions as they enter into college classrooms (Schneps & Sadler, 1988). The average of correct responses for all 27 TOAST items in this study was 43.16%. This result clearly indicates that the sample of undergraduate students were not proficient in the fundamental astronomy content standards covered by the TOAST. TOAST assessment for undergraduates, which we assume includes 40% (Lawrenz, Huffman, & Appeldoorn, 2005) future teachers, indicates that they are lacking in sufficient content knowledge in the area of

astronomy as prescribed by the American Astronomical Society (Partridge & Greenstein, 2003), the American Association for the Advancement of Science (1993), the National Science Education Standards (National Research Council Assessment, 1995) and the current Next Generation Science Standards (Lead States, 2013).

At the same time, results for most items reflect the body of literature on students' astronomical thinking. In areas with the largest extant research (e.g., lunar phases, season) the relationship between item results and the literature are quite strong. In cases where the research field is sparse in its investigation of students' mental models, such relationships are more difficult to assert. However, in no case do the results contradict the existing research. Additional research describing student thinking in qualitative terms may well remedy any shortfalls. Moreover, such research could be used to update and improve the TOAST and other assessment instruments for use in astronomy education and research.

In the interim, the TOAST is primarily intended to be used as a measure of students' mastery of the core concepts associated with an introductory astronomy course. In that capacity, its relationship to the extant research suggests it may be used as a tool to inform and improve instruction in an individual instructor's course, by administering it in a pre-post assessment manner. Administering it as a pre-assessment would allow the instructor to gain information about students' prior knowledge across the entire content of an ASTRO 101 course (Angelo & Cross, 1993). As education research literature across many fields indicates that an awareness of and response to students' prior knowledge is critical for effective instruction (Bransford, 1999). In administering the TOAST after instruction, the instructor can collect evidence of the effectiveness of instruction within their own classrooms.

The TOAST would also be appropriate to use as a tool to conduct astronomy education research and classroom level research, when employed across sections or semesters. This would allow the instructor or the researcher to compare the effectiveness of different instructional techniques or strategies. As a variation the TOAST can be used to measure the impact of adjusting instruction in one of the criteria "meta-categories." For instance, an instructor might note that students' understanding of patterns in the visible sky does not seem to improve after lecture-based or lecture and discussion-based instruction, and may decide to add a laboratory component for that portion of the course. The TOAST contains a sufficient number of items on this topic to provide for a sub-scale.

As a caution, note that comparing instructional techniques for a single instructor, or a coherent group of instructors, is different than comparing instructional techniques across many different instructors, especially when those instructors cannot be segregated into homogenous groups. We cannot speak to the usefulness of the TOAST, or any instrument, in these types of research studies. The presence of variables that have unintentionally been omitted from analysis and biasing factors are likely, and would skew results. As in all research, use of the TOAST should be done with a good dose of humility and understanding of unaccounted for variables, and the limits of any investigation into the complexities of teaching and learning.

AUTHOR INFORMATION

Stephanie J. Slater, Ph.D. is the Director of the CAPER Center for Astronomy & Physics Education Research. Stephanie@CAPERteam.com

Sharon Price Schleigh, Ed.D. is an Assistant Professor at East Carolina University.

Debra J. Stork, Ph.D. is a Professor of Education and Department Head of Teacher Education at the University of Dubuque.

REFERENCES

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. Oxford University Press.

Agan, L. (2004). Stellar ideas: Exploring students' understanding of stars. *Astronomy Education Review. 3(1),* 77-97.

Angelo, T. A., & Cross, K. P. (1993). Classroom assessment techniques: A handbook for faculty. *Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.*

Atwood, R. K., & Atwood, V. A. (1996). Preservice elementary teachers' conceptions of the causes of seasons. *Journal of research in science teaching*, *33*(5), 553-563.

Bailey, J. M. (2007). Development of a concept inventory to assess students' understanding and reasoning difficulties about the properties and formation of stars. *Astronomy Education Review*, *6*(2), 133-139.

Bardar, E. M., Prather, E. E., Brecher, K., & Slater, T. F. (2006). Development and validation of the Light and Spectroscopy Concept Inventory. *Astronomy Education Review*, *5*(2), 103-113.

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, *11*(5), 502-513.

Bisard, W. J., Aron, R. H., Francek, M. A., & Nelson, B. D. (1994). Assessing selected physical science and earth science misconceptions of middle school through university preservice teachers: Breaking the science misconception cycle. *Journal of College Science Teaching, 24*(1), 38-42.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington D.C.: National Academy Press.

Clark, D.B, Meneske, M., D'Angelo, C., Price Schleigh, S. & Ozdemir, G. (2014). Exploring sources of variation in studies of knowledge structure coherence: Comparing force meanings and force meaning consistency across two Turkish cities. *Science Education. 98*(1), 143-181.

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

Dai, M. F. W. (1991). Identification of misconceptions about the moon held by fifth and sixth graders in Taiwan and an application of teaching strategies for conceptual change. *Doctoral Dissertation, University of Georgia.*

Fanetti, T. M. (2001). *The relationships of scale concepts on college age students' misconceptions about the cause of the lunar phases* (Master's thesis, Iowa State University).

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309-333.

Hansson, L., & Redfors, A. (2006). "Swedish Upper Secondary Students' Views of the Origin and Development of the Universe," *Research in Science Education, 36*, 355.

Kikas, E. (2004). Teachers' conceptions and misconceptions concerning three natural phenomena. *Journal of Research in Science Teaching*, *41*(5), 432-448.

Lawrenz, F., Huffman, D., & Appeldoorn, K. (2005). Enhancing the instructional environment: Optimal learning in introductory science classes. *Journal of College Science Teaching*, *34*(7), 40.

Lead States, N. (2013). *Next generation science standards: For states, by states*. Washington, D.C.: National Academies Press.

Lindell, R. S., & Sommer, S. R. (2004). Using the lunar phases concept inventory to investigate college students' pre-instructional mental models of lunar phases. *AIP Conference Proceedings, 720*(1), 73-76. doi:10.1063/1.1807257

National Research Council Assessment. (1995). *National science education standards*. Washington D.C.: National Academies Press.

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

Osborne, R. J., & Gilbert, J. K. (1980). A method for investigating concept understanding in science. *European Journal of Science Education*, *2*(3), 311-321.

Partridge, B., & Greenstein, G. (2003). Goals for "ASTRO 101:" Report on workshops for department leaders. *Astronomy Education Review*, *2*(2), 46-89.

Philips, W. C. (1991). Earth science misconceptions. *Science Teacher*, *58*(2), 21-23.

Plummer, J. D. (2009). A cross-age study of children's knowledge of apparent celestial motion. *International Journal of Science Education*, *31*(12), 1571-1605.

Prather, E. E., Slater, T. F., & Offerdahl, E. G. (2002). Hints of a fundamental misconception in cosmology. *Astronomy Education Review*, *1*(2), 28-34.

Price Schleigh, S., Bosse, M., & Lee, T. (2011). Redefining curriculum integration and professional development: In-service teachers as agents of change. *Current Issues in Education, 14* (3), 12-pages.

Price Schleigh, S. (2015). Impacting society through astronomy undergraduate courses. American Physical Society annual conference, Invited Presentation H8.0001 vol. 60(4). Moving Astronomy Education Research into Teaching. April 2015. Baltimore MD.

Sadler, P.M., Coyle, H., Miller, J.L., Cook-Smith, N., Dussault, M., & Gould, R.R (2010). The astronomy and space concept inventory: Development and validation of assessment instruments aligned with K-12 National

Science Standards. *Astronomy Education Review, 8*(010111)

Sadler, P. M. (1992). The initial knowledge state of high school astronomy students. *Doctoral dissertation, Harvard University.*

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.

Schleigh, S.P., Clark, D.B., & Menekse, M. (2015). Constructed-response as an alternative to interviews in conceptual change studies: Students' explanations of force. *International Journal of Education in Mathematics, Science and Technology, 3*(1), 14-36.

Schneps, M. H., & Sadler, P. (1988). A private universe (film). *Santa Monica, CA: Pyramid Film and Video*.

Skam, K. (1994). Determining misconceptions about astronomy. *Australian Science Teachers Journal*, *40*(3), 63-67.

Slater (Parker), S. (2006). The impact of a kinesthetic astronomy curriculum on high school students' conceptual understanding of the seasons. *Masters' Capstone Project. Montana State University. Bozeman. MT.*

Slater, S. (2009, January). First results from the Test Of Astronomy STandards (TOAST) assessment instrument. In *Bulletin of the American Astronomical Society* (Vol. 41, p. 493).

Slater, S.J. (2014). Design and Development of the Test Of Astronomy STandards Conceptual Survey. *Journal of Astronomy & Earth Sciences Education, 1*(1), 1-22.

Slater, S., & Morrow, C. A. (2010, January). The Impact of a Kinesthetic Approach to Teaching Earth's Seasons. In *Bulletin of the American Astronomical Society* (Vol. 42, p. 555).

Slater, S. J., Morrow, C. A., & Slater, T. F. (2008). The impact of a kinesthetic astronomy curriculum on the content knowledge of at-risk students. In *Proceedings of the National Association for Research in Science Teaching, Baltimore, MD*.

Slater, S. J., Slater, T. F., Heyer, I., & Bailey, J. M. (2015). *Discipline-Based Education Research: A Scientist's Guide, 2nd edition*. Hilo, Hawai'i: Pono Publishing

Slater, S. J., Slater, T. F., Heyer, I., & Bailey, J. M. (2015). *Conducting Astronomy Education Research, 2nd edition*. Hilo, Hawai'i: Pono Publishing

Slater, S. J., Slater, T. F., & Shaner, A. (2008). Impact of backwards faded scaffolding in an astronomy course for pre-service elementary teachers based on inquiry. *Journal of Geoscience Education*, *56*(5), 408.

Slater, T. F. (1993). The effectiveness of a constructivist epistemological approach to the astronomy education of elementary and middle level in-service teachers. *Ph.D. Dissertation: University of South Carolina.*

Slater, T. F., Carpenter, J. R., & Safko, J. L. (1996). Dynamics of a constructivist astronomy course for in-service teachers. *Journal of Geoscience Education, 44*(6), 523-528.

Slater, T. F., Hufnagel, B., & Adams, J. P. (1999, May). Validating the Astronomy Diagnostics Test for Undergraduate Non-Science Majors. In *Bulletin of the American Astronomical Society* (Vol. 31, p. 937).

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Merrill.

Treagust, D. F., & Smith, C. L. (1989). Secondary students' understanding of gravity and the motion of planets. *School Science and Mathematics*, *89*(5), 380-391.

Turkoglu, O., Ornek, F., Gokdere, M., Suleymanoglu, N., & Orbay, M. (2009). On pre-service science teachers' preexisting knowledge levels about basic astronomy concepts. *International Journal of Physical Sciences*, *4*(11), 734-739.

Varma, S. (2006). *Preliminary item statistics using point-biserial correlation and p-values*. Educational Data Systems Inc.: Morgan Hill CA.

Vassar, M., & Crosby, J. W. (2008). A reliability generalization study of coefficient alpha for the UCLA Loneliness Scale. *Journal of Personality Assessment*, *90*(6), 601-607.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction*, *4*(1), 45-69.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535-585.

Wallace, C.S. (2011). An investigation into introductory astronomy students' difficulties with cosmology, and the development, validation, and efficacy of a new suite of cosmology lecture-tutorials. *Ph.D. Dissertation, University of Colorado.*

Wallace, C. S., & Bailey, J. M. (2010). Do concept inventories actually measure anything? *Astronomy Education Review*, *9*(1), 010116.

Zeilik, M. (2002). Birth of the Astronomy Diagnostic Test: Prototest evolution. *Astronomy Education Review*, *1*(2), 46-52.

NOTES