

Changing epistemological beliefs with nature of science implementations

Keith Johnson* and Shannon Willoughby

Montana State University, Bozeman, Montana 59717-3840, USA



(Received 22 June 2017; published 26 February 2018)

This article discusses our investigation regarding nature of science (NOS) implementations and epistemological beliefs within an undergraduate introductory astronomy course. The five year study consists of two years of baseline data in which no explicit use of NOS material was implemented, then three years of subsequent data in which specific NOS material was integrated into the classroom. Our original study covered two years of baseline data and one year of treatment data. Two additional years of treatment course data have revealed intriguing new insights into our students' epistemic belief structure. To monitor the evolution of belief structures across each semester we used student pre-post data on the Epistemological Beliefs About the Physical Sciences (EBAPS) assessment. The collected data were also partitioned and analyzed according to the following variables: college (Letters of Science, Business, Education, etc.), degree (BA or BS), status (freshman, sophomore, etc.), and gender (male or female). We find that the treatment course no longer undergoes significant overall epistemic deterioration after a semester of instruction. We also acquire a more detailed analysis of these findings utilizing the aforementioned variables. Most notably, we see that this intervention had a pronounced positive impact on males and on students within the college of Education, Arts & Architecture, and those with no concentration. Lastly, whether or not students believe their ability to learn science is innate or malleable did not seem to change, remaining a rigid construct with student epistemologies.

DOI: [10.1103/PhysRevPhysEducRes.14.010110](https://doi.org/10.1103/PhysRevPhysEducRes.14.010110)

I. INTRODUCTION

Epistemologies may be defined as a “set of views about the nature of knowledge, knowing, and learning [1]”. In general, there are two fields of thought regarding epistemologies. The first stance is that these views progress in stages, such as was pioneered by Perry and further supported in other work [2–4]. Here, a student may begin with the mindset that knowledge is inherited from authority and eventually come to the position that knowledge is constructed. Alternatively, Schommer provided compelling evidence that there exist multiple dimensions of epistemologies and that an individual has a varying degree of sophistication along each dimension, with similar work to follow this philosophy [5–7]. These dimensions have included constructs such as the tentativeness of knowledge and the innateness of ability.

More recent work has been based around epistemic beliefs being context dependent as well, that student epistemologies can be domain specific (physics, history,

mathematics, etc.) and even situation specific [8]. Our research, and thus our presentation of data with this paper, has adopted these latter philosophies for which epistemological beliefs are context dependent (domain, specifically) and exist along nonorthogonal dimensions.

In early 2012 Duncan published a paper discussing student epistemological beliefs within an introductory astronomy class [9]. His findings led him to conclude that basic inclusions of material involving the nature of science (NOS) within the classroom could have a measurable impact on students epistemic belief structures. Briefly, the nature of science may be thought of as “[...] the values and assumptions inherent to the development of scientific knowledge” [10]. Duncan’s claim is of value because, across the country, when undergraduates take a course in science their belief structures tend to deteriorate [6,9,11]. In other words, students’ views about knowledge and learning within science become less sophisticated after taking an introductory course in science. Duncan acknowledged shortcomings within his study and put a call out to others to investigate the influence of explicit NOS materials on epistemological beliefs within the classroom. Further details of the extant literature can be found in Ref. [12]. We proceed with a discussion of the methodology and any complications therein. Results within the groups of the study are subsequently discussed, as is a brief between-group comparison utilizing normalized change. Lastly, a

*keith.johnson12@montana.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International license*. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

context-dependent interpretation of effect size is presented, followed by a summary of the findings in this study.

II. METHODS

A. Design

In the fall of 2012 we began to investigate Duncan's claim with the collection of data from a large enrollment undergraduate introductory astronomy class (Astronomy 110). The course instructor (one of the authors) is well versed in physics education research and maintained an active classroom, frequently utilizing methods such as iClickers, group activities, and exams with a collaborative group portion. The instructor as well as the active classroom environment have remained constant throughout all years of this study.

The conclusion of the study occurred in spring 2017, having collected two years of baseline (control) course data from Fall 2012 to Spring 2014 and three years of modified (treatment) course data from Fall 2014 to Spring 2017. This large volume of data allowed us the ability to effectively explore several variables. These variables include gender, college (Letters of Science, Business, Education, etc.), degree (BA or BS), and status (freshman, sophomore, etc.). We chose to explore these variables as they were readily available and because other epistemological research has placed focus here [13–19]. Furthermore, although we focus on the claims made by Duncan, we are also motivated to explore the status quo of epistemic beliefs in the context of an introductory astronomy course.

Fundamentally, we sought to answer the question “What is the state of student epistemic beliefs within our introductory astronomy course and can basic course modifications, focused on NOS, prevent decay of student epistemologies towards science?” This overarching question branched into a subset of five research questions:

1. How do students' epistemic beliefs about science change over the course of a semester, as compared to our baseline (control) course?

2. How do students' epistemic beliefs about the sciences change after some basic course modifications and how does this compare to the changes seen in the baseline course?

3. Are there differences in epistemic beliefs when considering gender and do course modifications effect students of either gender differently?

4. Considering the college, degree, and status variables individually, are students within the baseline course more or less susceptible to epistemic change?

5. How do nature of science course modifications effect the epistemologies of students within these groups (college, degree, and status) differently?

Precise information regarding material and philosophies implemented within the modified course can be found in our original paper [12]. Within the modified course nature of science material was included at least weekly. This

allowed students plenty of practice with the material, so as to improve understanding of the content [10,20]. Implementation was not lecture based, but focused on group activities, individual exploration, and personal reflection. Students were frequently asked to interact with each other regarding course content.

Nature of science encompasses a wide range of topics which we narrowed down to the following: asking students to apply the scientific method, metacognitive tasks, model development, discussing the role of skepticism in scientific discovery, and finding connections to science in their daily lives. The new material was added to the course in lectures that had not previously filled the entire time allotted to each class meeting. These simple implementations were done with the hope that, if our findings were significant, instructors would not have to restructure an entire course in order to have a similar effect.

B. Population

As the population can be largely influential in conclusions that are made within any study, we discuss demographics. This study was conducted at a medium-sized, midwestern, land-grant institution with an open enrollment policy. Student body profiles at this university for Fall 2016 are as follows: 45% female, 55% male, 62% in-state residents, 34% out-of-state residents, and 4% international students. Of all the students, 83% identified as Caucasian, with the next largest ethnicity being Hispanic or Latino at roughly 4%. The remaining 13% of the student body consists predominantly of American Indian, Asian, African American, foreign students, and students of mixed race. Over the course of this study (Fall 2012 to Spring 2017), these profiles have remained approximately constant. Although these percentages represent the university as a whole, the instructor participating in the study thinks these values to be an accurate representation of the Astronomy 110 course population. As we analyze this system of student epistemic beliefs, it is important to note that any findings discussed in this paper are a representation of the demographics above.

C. Instrument

To measure student epistemic belief structures we relied on the epistemological beliefs about the physical sciences (EBAPS) assessment. This 30 question forced-choice, Likert-type instrument was developed and validated by Elby *et al.* The EBAPS measures epistemological beliefs along five nonorthogonal axes (structure of scientific knowledge, nature of knowing and learning, real-life applicability, evolving knowledge, and source of ability to learn), as well as an overall axis. Details regarding the axes of the EBAPS can be found in Table I, while more information regarding the EBAPS as a whole can be found at its host website [21,22]. Each axis measures student epistemological sophistication on a numeric scale from

TABLE I. Description of EBAPS axes.

-
-
1. *Structure of scientific knowledge.* Is [science] a coherent, conceptual, highly structured, unified whole? Expert views along this axis align with a view that the sciences are indeed coherent and unified.
 2. *Nature of knowing and learning.* Does learning science consist mainly of absorbing information? Or, does it rely crucially on constructing one's own understanding by working through the material actively, by relating new material to prior experiences, intuitions, and knowledge, and by reflecting on and monitoring one's own understanding? Thus, this axis gauges metacognition. Most experts agree that new information must be built into the web of our preexisting knowledge.
 3. *Real-life applicability.* Are scientific knowledge and scientific ways of thinking applicable only to restricted spheres such as the classroom or the laboratory? ... These items tease out students' views of the applicability of scientific knowledge as distinct from the student's own desire to apply science to real life, which depends on the student's interests, goals, and other nonepistemological factors. Professional scientists realize that thinking scientifically can be useful in everyday situations.
 4. *Evolving knowledge.* This dimension probes the extent to which students navigate between the twin perils of absolutism and extreme relativism. Experts avoid either peril. A score of four on this axis represents not falling into either trap. Other authors suggest this axis can also be viewed as measuring the extent to which students think that science is either tentative or settled. [23,24]
 5. *Source of ability to learn.* Is being good at science mostly a matter of fixed natural ability? Or, can most people become better at learning (and doing) science? As much as possible, these items probe students' epistemological views about the efficacy of hard work and good study strategies, as distinct from their self confidence and other beliefs about themselves. Education researchers typically agree that anyone can learn science and that it is not a fixed ability.
-
-

zero to four. On this scale, a zero represents a novicelike view whereas a four represents an expertlike view, with a gradient of sophistication between these values.

D. Procedure

The initial set of data in the study is the control, or the baseline course. This course represents data taken in both fall and spring semesters from Fall 2012 through Spring 2014, in which no explicit NOS material was incorporated. The second data set is that of the treatment group, what we call the modified course. The modified course represents data taken in the fall and spring semesters from Fall 2014 to Spring 2017, where focused NOS material was incorporated. Each semester numbered roughly 400 introductory astronomy students across two course sections.

The EBAPS was administered twice each semester. The pretest was given during the first week of class while the post-test was given during the final week of class. Each student received a physical copy of the EBAPS instrument and a bubble sheet scantron on which to mark their answers. They were given 20 min towards the end of class to complete the assessment.

A voluntary survey, the vast majority of students present when the assessment was given would still participate. Average student participation numbered around 350 students across two sections. After scantrons were collected, they were parsed to check for anomalies. Any scantrons that were unfinished were discarded, as were any that displayed obvious repetition (e.g., sheets which contained only "c" responses). Our study poses minimal risk to the student and as such has IRB exempt status.

From here, pre- and post-tests were compared side by side. Students who participated in both the EBAPS pretest and post-test were matched and kept, students who did not were discarded. This helps ensure a true representation of

changes in student beliefs across a semester. After this filtering process there were typically between 250 and 300 sets of matched data each semester.

Student response data were then transformed into relevant EBAPS information [21]. This transformation yields a numeric value (0–4) for every student along each of the five axes and an overall axis. Again, these axes are described in Table I.

To gauge the evolution of student epistemological beliefs across a semester, Wilcoxin signed-rank tests were used on each of the EBAPS axes. Calculations of effect size were done using the Cohen's d method with pooled standard deviations. Typical effect sizes are calculated by taking the difference in the mean of two populations and dividing by the control group standard deviation, essentially providing a signal-to-noise ratio as shown:

$$\text{Effect size} = \frac{\bar{X}_{\text{treatment}} - \bar{X}_{\text{control}}}{s_{\text{control}}}. \quad (1)$$

In place of the control standard deviation, we have opted to use a pooled standard deviation to help account for any noise that may have arisen in the modified group that may not have been present in our baseline:

$$s_{\text{pooled}} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}}. \quad (2)$$

Within this pooled standard deviation the population sizes for each group being compared, N_1 and N_2 , are taken into consideration alongside their standard deviations s_1 and s_2 , respectively. Unless otherwise noted, comparisons made in this paper would have $N_1 = N_2$, as they represent matched students between pretest and post-test data. When interpreting effect size a value of 0.2 may be considered a

small effect, 0.5 as medium, and 0.8 as large [25]. A thorough discussion of effect sizes is presented in works by Coe, Kirk, and Rice [26–28]. We have reason to believe that effect sizes of $|d| = 0.3$ could represent “large” effects in the context of this study, see the section “Assigning Meaning to Effect Size” within this paper for further insight.

Computations in this paper were run predominantly on IBM SPSS statistical analysis software. This software allowed for simple work with this study’s numerous variables and also provided a well-known statistical template from which others may conduct similar work.

III. COMPLICATIONS

A. Student background

This study has not dealt with several factors which have the potential to influence initial and evolving student epistemic beliefs [29,30]. Socioeconomic factors such as parental income, education, and occupation, or even individual factors such as the number of science courses previously taken by the student are not accounted for within this study. Student performance within a course has also been linked to instructor epistemologies, however, as the same instructor was present throughout this study we do not need to deeply consider variations in data regarding a change in authority figure [31,32].

B. Fall modified data

Ideally, pretest performance between all study semesters should not yield significant deviations from each other [11]. To test this, a one-way analysis of variance (ANOVA) was performed across all semesters within the study on EBAPS pretest scores (homogeneity of variances was upheld, an ANOVA assumption). Findings revealed statistically significant differences along the overall axis [$F(9, 2325) = 2.861, p = 0.002$]. Descriptives for the overall axis can be found in the Appendix.

A Duncan *post hoc* follow-up revealed that these significant differences were stemming predominantly from two of the fall semesters in the modified course (F15 and F16). Further probing with a Waller-Duncan *post hoc* Bayesian approach revealed that F14 may also be considered as contributing to this discrepancy. A plot of means by semester for the overall axis as well as the initial Duncan *post hoc* follow-up can be found in Fig. 1 and in the Appendix, respectively. The error bars in Fig. 1 represent 95% confidence levels.

Note from Table XIII in the Appendix that none of the spring semesters in the modified course differ significantly from those in the baseline course. The evidence seems to point towards an effect present between these modified course semesters: there are notably higher EBAPS scores in the fall as compared to the spring (and baseline). Hence, students in the fall semesters of the modified course do not

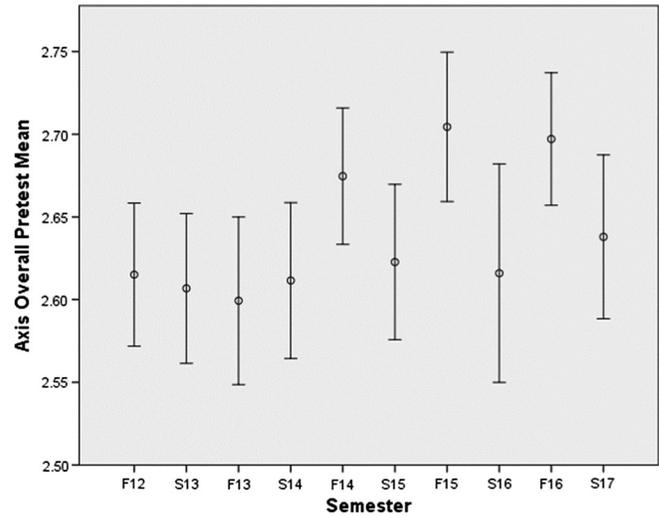


FIG. 1. Error bar plot of average pretest EBAPS score on overall axis by semester.

begin the course epistemologically the same as those in the baseline. As best we can tell there were no noteworthy changes in the pretest environment between fall and spring semesters in the modified course, relative to the baseline course. A meeting with our Director of University Studies and the Academic Advising Center also revealed no immediate probable cause for this trend. We are open to comments from the community regarding this finding.

Based on these results, if we separately group the modified course spring semester data (S15, S16, S17) and the modified course fall semester data (F14, F15, F16), a clearer analysis forms. Figure 2 shows a confidence interval plot (95% confidence) of the EBAPS overall axis mean by these groupings, baseline data were also portrayed for reference. Using guidelines from Cumming and Finch,

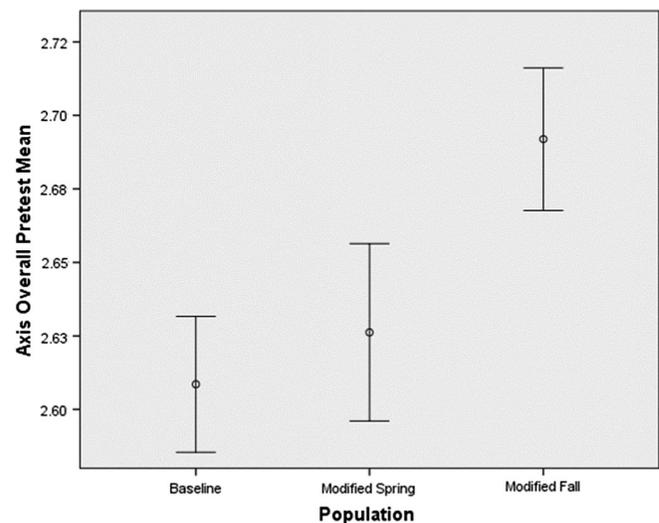


FIG. 2. Error bar plot of average pretest EBAPS score on overall axis with data combined.

confidence intervals dictate that fall semesters within the modified course come from different populations with $p < 0.01$ [33]. Since spring students in the modified course are epistemologically similar to those in the baseline, we believe the results of the grouped data from the spring portions of the modified course (S15, S16, S17) best convey the effect our changes had on the course. As such, the grouped spring semesters of the treatment course will be presented alongside baseline results throughout this paper.

Although the fall modified students will not be considered, we wish to incorporate their data deviations into our data analysis. This is done by considering the difference in mean pretest values for the EBAPS overall axis between the fall and the spring students of the modified course, then quantifying this difference in terms of effect size. We find that the modified fall students (F14, F15, F16) score higher on the overall EBAPS pretest axis than those in the spring with a Cohen's d effect of $d \approx 0.18$, a notably small effect [25] and one whose cause we have yet to definitively track down. As such we cannot assume that this is not just an effect due to noisy instrumentation. Thus, we move forward considering effect sizes around $|d| = 0.18$ as a discussion threshold. What we mean by this is that even if, say, Wilcoxin tests reveal significance ($p < 0.05$) along an EBAPS axis, the result will not be discussed unless their associated effect size is above $|d| = 0.18$. In truth, we will be aggressive with respect to our internal measure and use effect sizes at or above $|d| = 0.15$ with $p < 0.05$ as our significance criteria. To help emphasize the importance of effect size, we will favor the use of words such as "noticeable," "prominent," or "visible" when referring to findings exhibiting the aforementioned criteria, so as to delineate them from typical definitions of "significance" ($p < 0.05$).

IV. RESULTS WITHIN SEMESTER

A. Pre-post analysis: Total populations

All students within each data set were analyzed using a Wilcoxin related-samples t test and the results are shown below in Tables II–III. This analysis of the total

TABLE II. Baseline pretest and post-test results by EBAPS axis.^a

	Pretest	s.d.	Post-test	s.d.	Effect size	p value
Axis 1	2.25	0.46	2.28	0.49	0.06	0.04
Axis 2	2.60	0.52	2.52	0.56	−0.15	0.00
Axis 3	2.87	0.69	2.82	0.74	−0.06	0.23
Axis 4	2.60	0.78	2.52	0.77	−0.10	0.02
Axis 5	3.01	0.67	2.82	0.75	−0.27	0.00
Overall	2.61	0.35	2.54	0.41	−0.17	0.00

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

TABLE III. Modified spring pretest and post-test results by EBAPS axis.^a

	Pretest	s.d.	Post-test	s.d.	Effect size	p value
Axis 1	2.28	0.49	2.33	0.49	0.11	0.01
Axis 2	2.58	0.51	2.53	0.55	−0.10	0.03
Axis 3	2.84	0.73	2.89	0.74	−0.07	0.12
Axis 4	2.64	0.82	2.61	0.77	−0.03	0.36
Axis 5	3.06	0.68	2.95	0.70	−0.16	0.00
Overall	2.63	0.38	2.61	0.42	−0.03	0.77

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

population will yield the largest number of students for each data set within this paper and will thus yield the most statistically powerful conclusions present in this study. A bold axis label indicates prominence as previously defined ($|d| \geq 0.15$ with $p \leq 0.05$).

1. Baseline results: Consistent deterioration

Baseline results, involving 906 matched pre-post students, concur with those from other studies in that epistemological beliefs do indeed seem to deteriorate after a semester of instruction within a science course. In particular, we find noticeable losses along axis two (metacognition) and axis five (innate vs malleable mindset). In context, this means that after a semester of typical coursework the students are less likely to engage in metacognitive processes and are more likely to believe that one cannot become better at science through increased effort or a change in study behavior.

2. Modified spring results: Noteworthy improvement

Modified spring results (624 matched pre-post students in all) are encouraging, as overall epistemic beliefs are no longer undergoing a prominent deterioration. Axes two and five are not as pronounced as in the baseline, but do seem to be the most difficult to affect. Axis one (structure of scientific knowledge) hints at a positive response to the change, but the effect size present here does not clear our internal effect size noise filter.

B. Pre-post analysis: College

From now on, as data are separated, the relatively lower sample sizes will inherently lead to less statistical power, hence an increased likelihood that if significance exists it may not be detected. Fundamentally, this issue relates back to how confidence interval size is inversely related to sample size. Conclusions moving forward will consider and discuss this issue as it pertains to results.

Each particular data set may be separated by college. These colleges are Arts & Architecture (AA), Agriculture

(AG), Business (BU), Education/Health & Human Development (ED), Engineering (EN), Letters & Science (LS), and University College (UC). University College represents students who have no chosen concentration and cannot be grouped to a particular common program. Other colleges were present but lacked sufficient sample sizes, most numbering fewer than 20 students total across both baseline and modified data (e.g., College of Nursing; $N = 11$), and hence were not a part of the analysis.

TABLE IV. EBAPS pretest and post-test results for baseline by college.^a

College	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	p value	N
AA	Axis 1	2.33	0.41	2.36	0.49	0.06	0.34	172
	Axis 2	2.60	0.51	2.51	0.55	-0.16	0.06	
	Axis 3	2.91	0.68	2.95	0.67	0.05	0.60	
	Axis 4	2.72	0.75	2.54	0.79	-0.23	0.04	
	Axis 5	3.14	0.66	2.92	0.70	-0.33	0.00	
	Overall	2.67	0.33	2.60	0.42	-0.20	0.10	
BU	Axis 1	2.14	0.46	2.21	0.46	0.14	0.03	182
	Axis 2	2.54	0.52	2.45	0.52	-0.17	0.07	
	Axis 3	2.80	0.68	2.77	0.72	-0.05	0.72	
	Axis 4	2.47	0.79	2.50	0.75	0.03	0.66	
	Axis 5	3.00	0.66	2.78	0.76	-0.30	0.00	
	Overall	2.54	0.36	2.49	0.38	-0.12	0.10	
ED	Axis 1	2.21	0.42	2.22	0.45	0.02	0.62	113
	Axis 2	2.65	0.48	2.57	0.57	-0.14	0.09	
	Axis 3	2.76	0.60	2.65	0.74	-0.16	0.28	
	Axis 4	2.63	0.75	2.44	0.78	-0.24	0.03	
	Axis 5	2.89	0.67	2.64	0.78	-0.35	0.00	
	Overall	2.60	0.34	2.51	0.40	-0.24	0.02	
EN	Axis 1	2.31	0.52	2.40	0.44	0.17	0.37	68
	Axis 2	2.61	0.53	2.61	0.58	-0.01	0.84	
	Axis 3	2.99	0.88	2.97	0.72	-0.03	0.82	
	Axis 4	2.63	0.76	2.58	0.77	-0.06	0.85	
	Axis 5	3.07	0.72	2.94	0.83	-0.18	0.05	
	Overall	2.65	0.39	2.64	0.41	-0.02	0.55	
LS	Axis 1	2.30	0.49	2.36	0.53	0.11	0.08	202
	Axis 2	2.68	0.52	2.61	0.56	-0.13	0.12	
	Axis 3	2.98	0.69	2.92	0.78	-0.08	0.36	
	Axis 4	2.65	0.77	2.62	0.73	-0.04	0.80	
	Axis 5	3.00	0.66	2.90	0.71	-0.16	0.03	
	Overall	2.66	0.36	2.63	0.41	-0.09	0.17	
UC	Axis 1	2.22	0.44	2.17	0.46	-0.11	0.29	154
	Axis 2	2.51	0.52	2.40	0.56	-0.20	0.02	
	Axis 3	2.76	0.67	2.73	0.74	-0.05	0.75	
	Axis 4	2.50	0.82	2.42	0.76	-0.11	0.22	
	Axis 5	2.96	0.66	2.76	0.75	-0.29	0.00	
	Overall	2.54	0.33	2.44	0.39	-0.29	0.00	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

TABLE V. EBAPS pretest and post-test results for spring modified by college.^a

College	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	p value	N
AA	Axis 1	2.32	0.45	2.37	0.46	0.11	0.15	92
	Axis 2	2.55	0.48	2.42	0.60	-0.24	0.07	
	Axis 3	2.88	0.73	3.03	0.66	0.21	0.13	
	Axis 4	2.64	0.74	2.66	0.80	0.03	0.99	
	Axis 5	3.03	0.66	2.95	0.73	-0.12	0.28	
	Overall	2.63	0.36	2.61	0.42	-0.05	0.89	
BU	Axis 1	2.23	0.51	2.19	0.44	-0.08	0.49	113
	Axis 2	2.55	0.48	2.52	0.55	-0.07	0.47	
	Axis 3	2.77	0.73	2.73	0.73	-0.06	0.53	
	Axis 4	2.50	0.82	2.40	0.74	-0.12	0.26	
	Axis 5	3.03	0.70	2.87	0.64	-0.24	0.01	
	Overall	2.58	0.36	2.53	0.37	-0.15	0.09	
ED	Axis 1	2.27	0.47	2.36	0.51	0.19	0.04	81
	Axis 2	2.65	0.47	2.63	0.54	-0.06	0.65	
	Axis 3	2.79	0.69	2.77	0.76	-0.02	0.84	
	Axis 4	2.65	0.83	2.71	0.70	0.09	0.68	
	Axis 5	3.05	0.67	2.95	0.62	-0.14	0.13	
	Overall	2.63	0.37	2.64	0.42	0.03	0.52	
EN	Axis 1	2.38	0.44	2.47	0.52	0.19	0.10	81
	Axis 2	2.63	0.56	2.65	0.55	0.04	0.76	
	Axis 3	3.02	0.73	3.12	0.69	0.14	0.35	
	Axis 4	2.62	0.78	2.68	0.72	0.08	0.60	
	Axis 5	3.06	0.77	3.21	0.68	0.21	0.08	
	Overall	2.69	0.39	2.77	0.43	0.20	0.06	
LS	Axis 1	2.30	0.48	2.36	0.50	0.12	0.14	169
	Axis 2	2.56	0.51	2.50	0.53	-0.11	0.10	
	Axis 3	2.85	0.72	2.94	0.74	0.13	0.12	
	Axis 4	2.66	0.84	2.68	0.78	0.03	0.91	
	Axis 5	3.15	0.63	2.90	0.73	-0.36	0.00	
	Overall	2.65	0.38	2.61	0.42	-0.10	0.38	
UC	Axis 1	2.22	0.59	2.31	0.47	0.17	0.33	77
	Axis 2	2.55	0.57	2.50	0.55	-0.09	0.55	
	Axis 3	2.83	0.74	2.81	0.77	-0.02	1.00	
	Axis 4	2.74	0.87	2.50	0.83	-0.28	0.08	
	Axis 5	3.03	0.72	2.95	0.77	-0.10	0.55	
	Overall	2.59	0.45	2.59	0.42	-0.01	0.92	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

Wilcoxin test results separated by college are seen below for baseline (Table IV) and modified spring (Table V).

1. Results: Positive impact for AA, ED, EN, UC students

Across students of all colleges axis five (innate vs hard work) experiences prominent degradation from beginning to end of a semester within the baseline. The spring modified course is still hinting at detectable axis five decreases across essentially all colleges, BU ($p = 0.01$,

$d = -0.24$) and LS ($p < 0.01$, $d = -0.36$) with certainty. Namely, students in the modified course are now noticeably less likely, after instruction, to believe that innate ability dictates their learning, except those in BU and LS. Of great interest in the modified spring data is the engineering (EN) students, who not only have halted declines along axis five but now seem on the brink of positive prominence along this dimension ($p = 0.08$, $d = 0.21$).

Baseline AA ($p = 0.04$, $d = -0.23$), ED ($p = 0.03$, $d = -0.24$), and perhaps UC ($p = 0.22$, $d = -0.11$) students are most prone to experiencing a decrease along axis four (absolutism vs relativism). AA and ED modified course students no longer see noticeable decreases along axis four, yet UC students ($p = 0.08$, $d = -0.28$) seem unaffected and are thus still of concern. This means that AA and ED students are no longer prominently struggling with either an inability to delineate evidence-based statements from mere opinion, or the thought that all science is set in stone.

Axis three (applicability of science) remains firmly constant across all colleges in both the baseline and spring modified portions of the study. Thus, students have not undergone any prominent changes regarding their views on the applicability of science.

In the baseline course only UC students ($p = 0.02$, $d = -0.20$) experienced a prominent deterioration along axis two (metacognition), a deterioration which is not seen in the modified course. This finding indicates that the UC students are no longer notably worse at reflecting upon their own learning after a semester in Astronomy 110. It is worth noting there is reason to suspect that should larger sample sizes be present within each baseline college that axis two may actually be decaying noticeably (except with EN).

No college from the baseline data was seen to undergo visible change along axis one (structure of scientific knowledge). In the modified spring data ED students have undergone a noticeable increase along axis one ($p = 0.04$, $d = 0.19$), showing a greater ability to connect concepts within science. Outside of ED (which showed positive prominence) in the modified course, axis one seems to be on the verge of a visible increase across all other colleges except BU ($p = 0.49$, $d = -0.08$) and likely UC ($p = 0.33$, $d = 0.17$).

Overall, ED and UC students experience the greatest epistemological decay in the baseline, with AA students close behind. No colleges are observed to have prominent negative changes for the modified course. However, increased sample size for the modified course would likely see visible improvement overall for EN students ($p = 0.06$, $d = 0.20$), while revealing that BU students ($p = 0.09$, $d = -0.15$) may still undergo prominent overall negative change. With such low numbers in comparison it is difficult to state definitively but, in both baseline and modified spring, EN students appear to be least susceptible to overall

noticeable negative shifts in their epistemologies, an exception being baseline axis five ($p = 0.05$, $d = -0.18$). Concluding this section, we postulate that changes made to the course have been of greatest benefit to the overall epistemologies of AA, ED, and UC students.

C. Pre-post analysis: Degree

Students within the Astronomy 110 course were pursuing primarily either a Bachelor of Arts or a Bachelor of Science degree. As these were the only degrees with a sufficient sample size, the data was parsed accordingly. Bachelor of Fine Arts was the next largest degree, with only 87 total students across the 5 years of this study. Results and sample sizes regarding BA and BS students are found below in Table VI for baseline and Table VII for modified spring.

1. Results: Improvement for BS students

Baseline results indicate that BS students experience the most prominent epistemic decay, with axes two (metacognition), five (innate vs effort), and overall being the most influential. Axis four (absolutism vs relativism) may also be showing hints of trouble ($p = 0.08$, $d = -0.09$) while axis one (structure of scientific knowledge) appears promising ($p = 0.07$, $d = 0.06$). The BS modified spring students no longer experience any noteworthy overall decay such as that observed in the baseline course, but axis five losses persist. As seen in the baseline data, axis one is still promising ($p = 0.02$, $d = 0.12$) for this treatment data as well. To summarize, after instruction in the modified course, BS students are no longer more likely to

TABLE VI. EBAPS baseline pretest and post-test results by degree.^a

Degree	Axis	Pretest		Post-test		Effect		<i>N</i>
		s.d.		s.d.		size	<i>p</i> value	
BA	Axis 1	2.32	0.43	2.36	0.51	0.07	0.24	205
	Axis 2	2.63	0.48	2.58	0.56	-0.09	0.43	
	Axis 3	2.96	0.68	2.94	0.71	-0.03	0.78	
	Axis 4	2.66	0.73	2.61	0.75	-0.07	0.65	
	Axis 5	3.08	0.65	2.91	0.69	-0.26	0.00	
	Overall	2.67	0.33	2.62	0.40	-0.12	0.28	
BS	Axis 1	2.22	0.46	2.25	0.47	0.06	0.07	640
	Axis 2	2.58	0.53	2.49	0.56	-0.18	0.00	
	Axis 3	2.84	0.70	2.79	0.74	-0.07	0.27	
	Axis 4	2.55	0.79	2.48	0.77	-0.09	0.08	
	Axis 5	2.98	0.67	2.79	0.77	-0.28	0.00	
	Overall	2.58	0.36	2.52	0.40	-0.18	0.00	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

TABLE VII. EBAPS Modified spring pretest and post-test results by degree.^a

Degree	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	p value	N
BA	Axis 1	2.30	0.45	2.34	0.49	0.07	0.28	132
	Axis 2	2.53	0.53	2.46	0.57	-0.13	0.16	
	Axis 3	2.81	0.68	2.89	0.70	0.11	0.17	
	Axis 4	2.74	0.72	2.67	0.83	-0.09	0.37	
	Axis 5	3.05	0.64	2.90	0.72	-0.22	0.06	
	Overall	2.63	0.38	2.58	0.42	-0.11	0.31	
BS	Axis 1	2.27	0.50	2.33	0.49	0.12	0.02	449
	Axis 2	2.59	0.51	2.55	0.55	-0.08	0.13	
	Axis 3	2.85	0.74	2.87	0.75	0.03	0.61	
	Axis 4	2.60	0.83	2.58	0.75	-0.03	0.50	
	Axis 5	3.07	0.69	2.97	0.69	-0.15	0.01	
	Overall	2.63	0.39	2.62	0.42	-0.02	0.98	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

forgo metacognitive exercises but are still noticeably more likely to believe that scientific ability is innate.

Within the baseline, BA students undergo notable decreases along axis five ($p < 0.01$, $d = -0.26$) but no concerns otherwise. In general, the BA students in the modified course behave similarly to those in the baseline as there appears to still be issues with axis five ($p = 0.06$, $d = -0.22$), but no other near prominences to be concerned with. Thus, BA students remain essentially unaffected by changes made to the course.

D. Pre-post analysis: Gender

How male and female epistemologies begin and evolve within astronomy can be evaluated as well. Shown here are tables for baseline (Table VIII) and modified spring (Table IX), separated by gender. Student gender could not be determined for as much as 15% of each set of data and thus those data were not included.

1. Results: Deterioration for males more pronounced

Baseline males experience three prominent decay axes (two, four, and five) while males in the modified course experience no prominent occurrences of epistemological decay. We may safely state that males are noticeably more likely to, after a semester of unmodified instruction, either value opinion over evidence or think that science is uncompromising (axis four). They are also visibly less apt both to believe that hard work does improve their ability in science (axis two) and to engage in metacognitive learning opportunities (axis two). Males no longer suffer a noticeable decay from after the modified course along these axes.

TABLE VIII. EBAPS baseline pretest and post-test results by gender.^a

Gender	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	p value	N
Female	Axis 1	2.24	0.43	2.25	0.45	0.04	0.36	379
	Axis 2	2.65	0.50	2.58	0.52	-0.12	0.04	
	Axis 3	2.79	0.67	2.79	0.67	-0.01	0.92	
	Axis 4	2.63	0.75	2.57	0.74	-0.08	0.15	
	Axis 5	2.99	0.69	2.85	0.75	-0.20	0.00	
	Overall	2.61	0.34	2.57	0.37	-0.11	0.07	
Male	Axis 1	2.27	0.47	2.32	0.52	0.08	0.07	391
	Axis 2	2.57	0.53	2.48	0.57	-0.15	0.01	
	Axis 3	2.95	0.71	2.88	0.78	-0.08	0.27	
	Axis 4	2.62	0.79	2.50	0.77	-0.16	0.05	
	Axis 5	3.03	0.65	2.80	0.74	-0.33	0.00	
	Overall	2.62	0.37	2.54	0.43	-0.20	0.00	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

Females from the baseline course undergo only prominent epistemic change along axis five, where they are more likely to view scientific ability as set in stone. This notable deterioration along axis five for females ceases to exist in the modified course. Axis two (metacognitive ability) is on the verge of troublesome for females in both the baseline ($p = 0.04$, $d = -0.12$) and modified ($p = 0.03$, $d = -0.12$) study findings.

Axis five in the modified course an excellent example of statistical power. Neither males ($|d| = 0.11$, $p = 0.16$,

TABLE IX. EBAPS modified spring pretest and post-test results by gender.^a

Gender	Axis	Pretest	s.d.	Post-test	s.d.	effect size	p value	N
Female	Axis 1	2.29	0.47	2.35	0.49	0.12	0.07	259
	Axis 2	2.63	0.50	2.56	0.54	-0.12	0.03	
	Axis 3	2.77	0.73	2.84	0.74	0.09	0.21	
	Axis 4	2.69	0.80	2.68	0.74	-0.02	0.67	
	Axis 5	3.00	0.69	2.94	0.66	-0.10	0.16	
	Overall	2.63	0.38	2.62	0.41	-0.02	0.73	
Male	Axis 1	2.30	0.51	2.35	0.49	0.10	0.09	281
	Axis 2	2.55	0.51	2.53	0.58	-0.04	0.67	
	Axis 3	2.90	0.74	2.96	0.72	0.08	0.18	
	Axis 4	2.66	0.81	2.59	0.76	-0.09	0.23	
	Axis 5	3.08	0.69	3.00	0.73	-0.11	0.16	
	Overall	2.64	0.38	2.63	0.42	-0.02	0.59	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

$N = 281$) nor females ($|d| = 0.10$, $p = 0.16$, $N = 259$) experience noticeable change along axis five as defined in this paper ($|d| \geq 0.15$ with $p < 0.05$). Yet when combined, as seen in Table III, we discover a prominent change in modified spring students along axis five ($|d| = 0.16$, $p < 0.001$, $N = 624$). This also demonstrates the importance of total population results (Tables II and III) in comparison to the variable-based partitioned results within this, or any, paper.

In general, baseline data reveal that males are more prone to overall epistemological deterioration than females, although females do approach prominence ($p = 0.07$, $d = -0.11$). After a semester of instruction in the modified course, we no longer see either gender experiencing a prominent overall epistemic decay.

E. Pre-post analysis: Status

The final pre-post analysis within our study considers student status: freshman (FR), sophomore (SO), junior (JR), and senior (SR). Nearly all students could be linked with their status via online instructor-accessible campus information. The strongest cases here can be made for freshman and sophomores, as they comprised the majority of the introductory astronomy class. Results tables for baseline (Table X) and modified spring (Table XI) are displayed.

1. Results: Overall benefits for freshman

Baseline data show axis five (innate vs hard work) undergoing visible deterioration across all statuses with the exception of seniors, although that is a concerning axis ($p = 0.11$, $d = -0.14$) for them. These prominent declines along axis five ($p < 0.01$, $d = -0.26$) still occur for freshman in the modified spring data, but no longer occur for sophomores or juniors, and seniors ($p = 0.35$, $d = 0.14$) are no longer at risk. This means sophomores and juniors are visibly less prone to view their scientific knowledge as constant after the modified instruction, yet freshman still exhibit this.

Juniors ($p < 0.01$, $d = -0.35$) in the baseline were the only class to show noticeable decay for axis four (absolutism vs relativism), but ceased this decay in the modified course. One might suggest that the negative baseline change is due to a higher count of males than females for juniors who enroll in Astronomy 110. An independent Wilcoxin test revealed that JR males ($p = 0.03$, $d = -0.42$, $N = 47$) showed prominence as compared to the females ($p = 0.14$, $d = -0.23$, $N = 62$) for this baseline data. Yet with greater numbers we expect females would also exhibit significance; regardless, JR males do seem to be more prone to decay along axis four in baseline data. No other status underwent notable change with axis four in either the baseline or modified data. We thus find that juniors in the modified course no longer struggle notably, or at all, with either the idea that all scientific

TABLE X. EBAPS baseline pretest and post-test results by status.^a

Status	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	p value	N
FR	Axis 1	2.18	0.43	2.19	0.44	0.01	0.91	374
	Axis 2	2.52	0.52	2.41	0.53	-0.23	0.00	
	Axis 3	2.83	0.68	2.77	0.74	-0.08	0.28	
	Axis 4	2.50	0.79	2.46	0.77	-0.05	0.57	
	Axis 5	3.05	0.66	2.86	0.73	-0.28	0.00	
	Overall	2.56	0.34	2.48	0.37	-0.22	0.00	
SO	Axis 1	2.28	0.47	2.35	0.49	0.15	0.00	294
	Axis 2	2.64	0.49	2.58	0.58	-0.12	0.13	
	Axis 3	2.88	0.68	2.87	0.72	-0.02	0.84	
	Axis 4	2.62	0.77	2.57	0.76	-0.07	0.33	
	Axis 5	2.98	0.70	2.78	0.76	-0.28	0.00	
	Overall	2.63	0.36	2.58	0.42	-0.13	0.08	
JR	Axis 1	2.33	0.48	2.29	0.49	-0.10	0.51	125
	Axis 2	2.63	0.53	2.61	0.52	-0.04	0.73	
	Axis 3	2.85	0.68	2.77	0.71	-0.11	0.40	
	Axis 4	2.73	0.72	2.47	0.72	-0.35	0.00	
	Axis 5	3.03	0.59	2.79	0.79	-0.34	0.00	
	Overall	2.65	0.31	2.56	0.39	-0.25	0.01	
SR	Axis 1	2.30	0.48	2.38	0.57	0.14	0.09	111
	Axis 2	2.67	0.56	2.62	0.58	-0.08	0.22	
	Axis 3	2.98	0.74	2.93	0.78	-0.06	0.83	
	Axis 4	2.73	0.79	2.65	0.81	-0.09	0.58	
	Axis 5	2.95	0.71	2.84	0.76	-0.14	0.11	
	Overall	2.66	0.41	2.64	0.47	-0.05	0.58	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

knowledge is relative nor the belief that this knowledge is unchanging.

As seen throughout this study, axis three (applicability of science) undergoes no visible change in either the baseline or modified portion of the study. The exception to this finding is sophomores in the modified course, who experienced prominent positive change ($p < 0.01$, $d = 0.25$). Hence, after a semester of instruction in the modified course, sophomores no longer have notable decay in their ability to make connects to science in their everyday lives.

Axis two (metacognition) sees prominent deterioration for freshman ($p < 0.01$, $d = -0.23$) and only freshman within the baseline. Therefore, these freshman are noticeably less likely to utilize metacognition after a semester in the baseline course. Unfortunately, this is an aspect of freshman epistemologies which does not change for the modified course ($p < 0.01$, $d = -0.18$). Sophomores remained unaffected along axis two in the modified course while both juniors ($p = 0.31$, $d = 0.17$) and seniors ($p < 0.28$, $d = 0.20$) have strong, but not significant, positive increases along this dimension.

TABLE XI. EBAPS modified spring pretest and post-test results by status.^a

Status	Axis	Pretest	s.d.	Post-test	s.d.	Effect size	<i>p</i> value	<i>N</i>
FR	Axis 1	2.24	0.48	2.29	0.49	0.10	0.09	338
	Axis 2	2.56	0.49	2.46	0.55	-0.18	0.00	
	Axis 3	2.79	0.73	2.76	0.73	-0.04	0.59	
	Axis 4	2.57	0.85	2.50	0.79	-0.08	0.21	
	Axis 5	3.03	0.68	2.85	0.73	-0.26	0.00	
	Overall	2.59	0.36	2.54	0.41	-0.13	0.06	
SO	Axis 1	2.32	0.50	2.38	0.46	0.12	0.08	187
	Axis 2	2.61	0.50	2.55	0.55	-0.11	0.41	
	Axis 3	2.87	0.73	3.05	0.70	0.25	0.00	
	Axis 4	2.71	0.77	2.72	0.75	0.01	1.00	
	Axis 5	3.09	0.70	3.04	0.65	-0.07	0.24	
	Overall	2.66	0.40	2.67	0.39	0.03	0.45	
JR	Axis 1	2.40	0.45	2.41	0.49	0.04	0.70	58
	Axis 2	2.61	0.53	2.70	0.51	0.17	0.31	
	Axis 3	2.89	0.73	3.10	0.64	0.29	0.07	
	Axis 4	2.72	0.75	2.80	0.66	0.11	0.48	
	Axis 5	3.14	0.62	3.15	0.68	0.02	0.34	
	Overall	2.71	0.36	2.77	0.37	0.15	0.12	
SR	Axis 1	2.30	0.52	2.40	0.60	0.16	0.25	41
	Axis 2	2.54	0.65	2.66	0.57	0.20	0.28	
	Axis 3	3.02	0.72	2.93	0.86	-0.12	0.43	
	Axis 4	2.74	0.78	2.69	0.74	-0.07	0.71	
	Axis 5	3.08	0.74	3.11	0.65	0.05	0.61	
	Overall	2.64	0.48	2.71	0.48	0.14	0.35	

^aScores range from 0 to 4. Bold indicates prominence with $|d| \geq 0.15$ and $p \leq 0.05$. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

Interestingly, axis one (structure of scientific knowledge) has a prominent increase for baseline sophomores ($p < 0.01$, $d = 0.15$) and is trending that way for seniors ($p = 0.09$, $d = 0.14$). Modified data show no significant change along axis one for any status. The significant increase for SO students along axis one as seen in the baseline is likely still present in modified spring data ($p = 0.08$, $d = 0.12$); however, the lower sample size (hence lower statistical power) simply is not allowing this effect to be detected. In the end, the implementations put into the modified course cause no real change compared to what was observed in the baseline for the structure of scientific knowledge.

In effectively no way do baseline seniors experience declines in overall epistemic beliefs. Meanwhile, freshman and juniors in the baseline course undergo visible overall epistemological decay and it is likely that sophomores ($p = 0.08$, $d = -0.13$) do as well. Although no overall epistemic deterioration is seen for any class in the modified course, it is reasonable to believe freshman still experience this ($p = 0.06$, $d = -0.13$).

The overarching trend in this data would seemingly be that multiple axes across all statuses other than freshman no longer undergo deterioration, and several now approach prominence with positive effect sizes (indicating improvement).

V. NORMALIZED CHANGE

To assist in comparing performance between semesters, normalized change was utilized. Normalized change is a construct put forth by Marx and Cummings and is calculated in a similar manner as the average of gains but removes students who score alike extreme values (0 or 4, in the case of the EBAPS) on both pretest and post-test [34]. The precise method of calculation is as follows:

$$c = \begin{cases} \frac{\text{post-pre}}{100\%-\text{pre}} & \text{post} > \text{pre} \\ \text{remove} & \text{pre} = \text{post} = 0\%, 100\% \\ \frac{\text{post-pre}}{\text{pre}} & \text{post} < \text{pre}. \end{cases} \quad (3)$$

Normalized change between student pretest and post-test EBAPS scores were calculated for all axes of baseline, modified spring, and modified fall data. A Kruskal-Wallis test was then performed, comparing distributions of normalized change between the aforementioned populations. Descriptives of the populations in this comparison can be found in the Appendix.

Pairwise comparisons with adjusted p values revealed that mean normalized change values along axis five ($p = 0.007$) and the overall axis ($p = 0.001$) were significantly higher for the modified spring population compared to the baseline course. No other differences were found in mean normalized change between these populations. The shift in mean normalized change for modified spring data compared to baseline data is given by an effect size of $d = 0.17$ as measured, along the overall axis and $d = 0.13$ as measured along axis five. Thus the changes implemented to the course have resulted in noticeably improved overall epistemic beliefs, comparatively.

VI. ASSIGNING MEANING TO OUR EFFECT SIZES

Although Cohen provides qualifications of meaningful effect sizes, we find it more insightful to give effect sizes presented in this paper a context outside predetermined norms. We shall do this by simulating the epistemological growth of students as they progress from freshman to senior. Consider that the bulk of students at our institution often take the introductory astronomy course as one of only two required science electives within the core curriculum. Consequently, this means the vast majority of students that take Astronomy 110 do so having taken at most one previous collegiate science course. If one then considers the difference in average EBAPS pre-test scores between

freshman and seniors, a glimpse of epistemic growth across a collegiate career can be acquired and represented as an effect size.

Baseline data from Table X for EBAPS overall pre-test scores of freshman and seniors leads to an effect size (using pooled standard deviation) of $d = 0.28$. While by no means definitive, an effect size of $d = 0.28$ does give us a proxy for the epistemic growth of students at our institution over their college careers as quantified by the EBAPS.

The approach outlined above may also be used for essentially any of the variables within this study, but let us present an example using the variable “college” (EN, LS, BU, etc.). In SPSS we may split the college variable up by status and acquire the overall axis baseline EBAPS pretest data of freshman for BU ($M = 2.46$, $s = 0.30$, $N = 80$), ED ($M = 2.56$, $s = 0.34$, $N = 31$), and LS ($M = 2.65$, $s = 0.34$, $N = 63$) students. Senior numbers were lower than thought acceptable, and thus were combined with juniors. We were justified in doing so, as a Mann-Whitney U test of independence showed no significant differences between SR and JR for BU ($U = 297.5$, $p = 0.349$), ED ($U = 186.5$, $p = 0.700$), or LS ($U = 451.0$, $p = 0.211$) with respect to the overall axis in EBAPS baseline pretest scores. Relevant data of upperclassmen by BU ($M = 2.56$, $s = 0.41$, $N = 53$), ED ($M = 2.58$, $s = 0.39$, $N = 44$), and LS ($M = 2.74$, $s = 0.35$, $N = 66$) were then subsequently acquired. It follows that effect sizes indirectly representing collegiate epistemic growth for BU ($d = 0.29$), ED ($d = 0.05$), and LS ($d = 0.26$) were calculated.

The goal in this section has been to place the effect sizes presented in this paper into context by showing what typical effect sizes may be attributed to student epistemic growth across their collegiate careers. These previous effect sizes show that while $d = 0.28$ is perhaps a robust value for this course, an effect size as low as $d = 0.05$ (like that for ED students) could also represent the cumulative effect of a collegiate career on epistemic beliefs regarding general science. With respect to both this course, the instrument being used, and the context provided, we postulate that an effect size of $d = 0.3$ may be considered a “large” effect within our findings. There is, of course, room for error within this value as issues such as attrition by major are not being accounted for.

VII. SUMMARY

We now summarize these findings alongside our research questions, as well as other aspects of student epistemologies within this study.

A. Research questions

Recall that our original motivation for this work was based around the fundamental question: “What is the state

of student epistemic beliefs within our introductory astronomy course and can basic course modifications, focused on NOS, prevent decay of student epistemologies towards science?” This study was intended to supplement the sparse literature on introductory astronomy student epistemologies, as well as investigate the impact of NOS implementations. These two factors led us to address a set of research questions, the findings of which we now summarize.

1. How do students’ epistemic beliefs about science change over the course of a semester, as compared to our baseline (control) course?

The baseline portion of the study yielded significant decreases in epistemic beliefs along axes two, four, five, and overall, as measured by the EBAPS instrument. These data indicate that students were less apt to engage in metacognitive practices after a semester of instruction, relying more heavily on simply absorbing information (memorization). They were also either less capable of distinguishing opinion from evidence-based argument or more likely to believe all scientific findings are set in stone. Lastly, students leaving the course were more likely to believe that scientific ability is a fixed trait and not something that can be improved with hard work. Refer to Table II for the baseline data.

2. How do students’ epistemic beliefs about the sciences change after some basic course modifications and how does this compare to the changes seen in the baseline course?

When analyzing across semesters within the modified course we saw that the fall semester performed notably better on EBAPS pretest scores than their spring counterparts. The spring semesters of the modified course scored similarly to the baseline on the pretest and thus were, as measured by the EBAPS, epistemologically identical to the baseline. As such, only the spring semesters were compared to the baseline within this study. Data for the spring modified students can be seen in Table III.

The modified spring students responded positively toward the NOS materials as they ceased significant deterioration of their overall belief structures and even began moving toward significant increases in how they view the structure of scientific knowledge (axis one). This means our students may now be more inclined to see science as an interconnected weaving of ideas and concepts, as opposed to a collection of isolated “facts.” Declines in metacognitive ability (axis two) were not as prominent for the modified spring course, but certainly still appear to be an area of concern. The most troublesome issue after a semester of instruction is that students are still more likely to believe that their ability to learn within science is innate, as opposed to malleable.

3. Are there differences in epistemic beliefs when considering gender and do course modifications effect students of either gender differently?

Gender EBAPS information for the baseline and spring modified course can be found in Tables VIII and IX, respectively. Male students in the baseline course appear to be more prone to overall epistemological decay than females, noticeably so with metacognitive ability, evolving knowledge, and source of ability to learn (axes two, four, and five, respectively). Baseline females only experienced visible decay in that they are less likely to believe they can improve their scientific ability with hard work (axis five). After a semester of instruction in the baseline, males clearly stood apart from females in that males were more likely to either view scientific knowledge as unyielding or were less capable of distinguishing opinion from evidence-based arguments (axis four).

Modified spring data revealed no prominent deterioration for either sex along any axis, although their is good reason to believe views regarding axis five (innate vs malleable mindset) are still worrisome.

4. Considering the college, degree, and status variables individually, are students within the baseline course more or less susceptible to epistemic change?

Baseline data revealed noticeable deterioration along axis five across all college domains (Table IV). Arts & Architecture students as well as Education students undergo prominent decay along axis four (evolving knowledge), a trait that no other colleges exhibit. University Studies students are the only college category that experience noticeable losses in their metacognitive views (axis two).

Bachelor of Science degree pursuers experience overall epistemic decay, whereas Bachelor of Arts students do not. Both groups, however, do show noticeable losses along axis five (source of ability to learn). Of the two groups, Bachelor of Science students also experience a visible loss in their ability to engage in metacognitive processes after a semester of instruction (axis two). Table VI contains the relevant information regarding these two degrees.

Within the baseline course freshman underwent significant declines on axis two (metacognition), five (innate vs effort), and overall. Views regarding axis five also underwent visible deterioration for sophomores and juniors. Sophomores were the only group to experience a noticeable increase for their beliefs on the structure of scientific knowledge (axis one) and juniors were the only class to have a prominent decay in their ability to delineate opinion from evidence (axis four). The epistemic beliefs of seniors experience no significant change. These findings regarding student status are presented in Table X.

5. Do nature of science course modifications effect the epistemologies of students within these groups (college, degree, and status) differently?

Within the modified spring course (Table V), the EN students represented the only college to not only cease epistemic declines along axis five but then actually move toward significant improvement here. Overall, EN students seem to have the most resilient epistemologies, experiencing minor degradations in baseline and no degradation in the modified course. None of the colleges experienced essentially any change for better or worse along axis three throughout the baseline and modified portions of the study. AA and ED students had experienced the most trouble out of any of the colleges along axis four, however, the prominent decreases along this axis ceased to exist in the modified spring course. Furthermore, ED students in the modified spring course have undergone noticeable increases along axis one.

When partitioned by degree, we no longer see any overall epistemic decay for either BA or BS students (Table VII). BS students also no longer undergo notable losses along axis two (metacognition), but do still seem to suffer from deteriorated beliefs regarding their views on hard work (axis five).

The overall axis for freshman in the modified spring course no longer shows visible decay, yet axes two and five remain problematic. Sophomores originally had strong decays along axis five, this behavior ceased in the modified spring course. Furthermore, in both the modified and baseline course sophomores showed improvement along axis one and even had a prominent increase along axis three in the modified course. Juniors had visible decays for axes four, five, and overall in the baseline. No significant decays were found along any axes for juniors in the modified spring results. Seniors were firm in their epistemologies, showing no prominent increase or decrease along any axes in both the baseline and modified portions of this study. It should be noted that student numbers were lower for juniors and seniors in both portions of the study. These data may be found in Table XI.

B. Fall semesters within modified course

The discrepancy between spring and fall semesters within only the modified course remains a mystery. As such, we cannot be completely certain that the positive outcomes seen in the spring are attributable to solely the implemented NOS material, despite the spring population being epistemologically identical to our baseline population. Although not presented, course changes seemed to have little effect on the modified fall students. Nevertheless, what we can say is that the cause of this discrepancy does not seem to be linked to any of the variables put forth within this study. That is, the differences in the spring and fall do not seem to be caused by college, status, degree, gender, or the student proportionality within those variables between

populations. Whatever the cause, students in the fall modified course scoring higher on the pretest across effectively every axis and every primary variable (college, degree, gender, status). Independent probing of the data has found that while high-school GPA and whether or not a student has taken college preparatory courses are capable of generating the differences we see, these factors cannot be definitively stated to be the cause.

C. Conclusions

We believe the changes incorporated into the classroom, as compared to the baseline course, have the potential to create a non-negligible positive effect on student epistemologies, as seen with the epistemologically similar modified spring students. Those who benefited the most from the intervention appeared to be males. Also experiencing notable benefits were students in Education, Arts & Architecture, and those who were without a concentration (UC). Engineering students also deserve mention, as they no longer undergo negative change along any dimension and are now on the cusp of significant positive improvement for axes one (structure of scientific knowledge), three (real-life applicability), five (innate vs effort), and overall. In general, axis five has remained the most difficult to effect. We suspect it may have more to do with student views regarding study strategies (which was not a focus of the implemented material) than it does fixed ability; however, more work is needed to support this claim. Results from this study may be utilized to guide focused epistemological work for students of a particular gender, status, or college with respect to the EBAPS axes. The instructor for the course wishes to continue the use of the additional NOS material, as they have received favorable student feedback and finds the new material to make for a more compelling, engaging classroom.

VIII. FUTURE WORK

Continued analyses between and within the populations of our study will be pursued, both to further explore the variables involved in this study as well as to carry on the investigation of the anomalous spring or fall behavior in the modified course. We will also move towards analyses of individual question responses on the EBAPS and student interview findings.

ACKNOWLEDGMENTS

We would like to thank the Montana Space Grant Consortium for funding, Arthur Bangert for statistical consultation and Diane Donnelly, our Director of University Studies and the Academic Advising Center, for her insight.

APPENDIX: ADDITIONAL INFORMATION

See Tables XII–XIV.

TABLE XII. Pretest mean across all study semesters on EBAPS overall axis.^a

Semester	N	Pretest	s.d.
F12	231	2.61	0.33
S13	247	2.61	0.36
F13	194	2.60	0.36
S14	234	2.61	0.37
F14	273	2.67	0.35
S15	259	2.62	0.38
F15	264	2.70	0.37
S16	134	2.62	0.39
F16	267	2.70	0.33
S17	232	2.64	0.38

^aScores range from 0 to 4.

TABLE XIII. Duncan *post hoc* results on pretest EBAPS averages.^a

Semester	N	Group 1	Group 2
F13	194	2.60	
S13	247	2.60	
S14	234	2.61	
F12	231	2.62	
S16	134	2.62	
S15	259	2.62	
S17	232	2.64	2.64
F14	273	2.67	2.67
F16	267		2.70
F15	264		2.70
Sig.		0.059	0.075

^aScores range from 0 to 4.

TABLE XIV. Descriptives for normalized change by population.^a

	Population	N	Mean	s.d.
Axis 1	Baseline	906	0.032	0.23
	Modified spring	624	0.051	0.23
Axis 2	Baseline	906	0.018	0.27
	Modified spring	624	0.030	0.26
Axis 3	Baseline	890	0.102	0.40
	Modified spring	616	0.146	0.41
Axis 4	Baseline	895	0.041	0.41
	Modified spring	614	0.080	0.43
Axis 5	Baseline	882	0.036	0.39
	Modified spring	611	0.086	0.39
Overall	Baseline	906	0.005	0.18
	Modified spring	624	0.035	0.18

^aScores range from 0 to 4. Axis 1: Structure of scientific knowledge, Axis 2: Nature of knowing and learning, Axis 3: Real-life applicability, Axis 4: Evolving knowledge, Axis 5: Source of ability to learn.

- [1] A. Elby, Getting Started with Research on Epistemologies and Expectations, *Getting Started in PER*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2011), Vol. 2, pp. 1–33.
- [2] W. G. Perry, *Forms of Ethical and Intellectual Development in the College Years: A Scheme* (Rinehart and Winston, New York, 1970).
- [3] M. B. Magolda, *Knowing and Reasoning in College* (Jossey-Bass, San Francisco, 1992).
- [4] B. K. Hofer and P. R. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, *Rev. Educ. Res.* **67**, 88 (1997).
- [5] M. Schommer, Effects of beliefs about the nature of knowledge on comprehension, *J. Educ. Psychol.* **82**, 498 (1990).
- [6] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in Introductory Physics, *Am. J. Phys.* **66**, 212 (1998).
- [7] D. Hammer, Epistemological beliefs in introductory physics, *Cognit. Instr.* **12**, 151 (1994).
- [8] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, *J. Learn. Sci.* **12**, 53 (2003).
- [9] D. Duncan and L. Arthurs, Improving student attitudes about learning science and student scientific reasoning skills, *Astron. Educ. Rev.* **11**, 010102 (2012).
- [10] N. Lederman, Students' and teachers' conceptions of the nature of science: A review of the research, *J. Res. Sci. Teach.* **29**, 331 (1992).
- [11] W. K. Adams, K. K. Perkins, N. S. Podelfsky *et al.*, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
- [12] S. Willoughby and K. Johnson, Epistemic beliefs of non-STEM majors regarding the nature of science: Where they are and what we can do, *Am. J. Phys.* **85**, 461 (2017).
- [13] K. Hakan and E. Munire, Profiling individual differences in undergraduates epistemological beliefs: Gender, domain and grade differences, *Procedia Soc. Behav. Sci.* **31**, 738 (2012).
- [14] L. Ding and P. Zhang, Making of epistemologically sophisticated physics teachers: A cross-sequential study of epistemological progression from preservice to in-service teachers, *Phys. Rev. Phys. Educ. Res.* **12**, 020137 (2016).
- [15] S. P. Bates *et al.*, How attitudes and beliefs about physics change from high school to faculty, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020114 (2011).
- [16] M. Schommer, Effects of beliefs about the nature of knowledge on comprehension, *J. Educ. Psychol.* **82**, 498 (1990).
- [17] M. Schommer, Epistemological development and academic performance among secondary students, *J. Educ. Psychol.* **85**, 406 (1993).
- [18] M. B. Paulsen and C. T. Wells, Domain differences in the epistemological beliefs of college students, *Res. High. Educ.* **39**, 365 (1998).
- [19] B. King and S. Magun-Jackson, Epistemological beliefs of engineering students, *J. Technol. Stud.* **35**, 56 (2009).
- [20] D. Allchin, *Teaching the Nature of Science* (SHiPS Education Press, Saint Paul, MN, 2013), pp. 3–27.
- [21] The idea behind ebaps, <http://www2.physics.umd.edu/~elby/EBAPS/idea.htm>.
- [22] <http://www2.physics.umd.edu/~elby/EBAPS/home.htm>.
- [23] A. Elby, Helping physics students learn how to learn, *Phys. Educ. Res., Am. J. Phys., Suppl.* **69**, S54 (2000).
- [24] G. Duffy and S. Chance, Improving Engineering Students Design Skills in a Project-based Learning Course by Addressing Epistemological Issues, *Proceedings of the 40th SEFI Conference, Thessaloniki, Greece* (2012).
- [25] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, N.J., 1988).
- [26] R. Coe, It's the Effect Size, Stupid: What effect size is and why it is important, *Proceedings of the Annual Conference of the British Educational Research Association, University of Exeter, 2012*, <https://www.leeds.ac.uk/educol/documents/00002182.htm>.
- [27] R. Kirk, Practical significance: A concept whose time has come, *Educ. Psychol. Meas.* **56**, 746 (1996).
- [28] M. Rice and G. Harris, Comparing effect sizes in follow-up studies: ROC, Cohen's d and r , *Law Hum Behav* **29**, 615 (2005).
- [29] Z. Hazari, and R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, *Sci. Educ.* **91**, 847 (2007).
- [30] A. V. Maltese and R. H. Tai, Eyeballs in the Fridge: Sources of early interest in science, *Int. J. Sci. Educ.* **32**, 669 (2010).
- [31] M. D. Sundberg, M. L. Dini, and E. Li, Decreasing course content improves student comprehension of science and attitudes towards science in freshman biology, *J. Res. Sci. Teach.* **31**, 679 (1994).
- [32] A. Wayne and P. Youngs, Teacher characteristics and student achievement gains: A review, *Rev. Educ. Res.* **73**, 89 (2003).
- [33] G. Cumming and S. Finch, Inference by eye: Confidence intervals and how to read pictures of data, *Am. Psychol.* **60**, 170 (2005).
- [34] J. D. Marx and K. Cummings, Normalized change, *Am. J. Phys.* **75**, 87 (2007).