Contribution of self-directed, naked-eye observations to students' conceptual understanding and attitudes towards astronomy

David R. Gozzard[*] and Marjan G. Zadnik

Department of Physics, School of Physics, Mathematics & Computing,
The University of Western Australia, 35 Stirling Hwy, Crawley 6009, Western Australia

(Received 30 April 2020; revised 25 February 2021; accepted 2 April 2021; published 6 May 2021)

Introductory university astronomy courses have emerged as a powerful opportunity to improve science understanding, literacy, and appreciation of the scientific method to a wide range of future citizens and voters. Students in an introductory astronomy course were instructed to record and analyze their naked-eye astronomical observations over a semester (13 weeks). The effect this activity had on students' learning of basic astronomy concepts and attitudes towards astronomy and science were measured using an astronomy concept diagnostic test and attitudes survey administered at the start and at the end of the semester, and compared with a similar introductory astronomy course that did not include the observing assignment. The results suggest that the observing diary is a positive learning experience for the majority of students, and that completing this assessment assists in developing students' deeper engagement with astronomy and astronomy concepts. We recommend the use and assessment of observing diaries as an effective learning activity in introductory astronomy classes.

DOI: 10.1103/PhysRevPhysEducRes.17.010134

I. INTRODUCTION

For tertiary students not majoring in science, technology, engineering, and mathematics (STEM) degrees, the inclusion of units of science in a general education curriculum is seen as a way to teach important critical thinking skills, as well as to tackle issues such as science and information literacy, sustainable development and responsible citizenship, and interest in and support for science by future citizens and voters [1–7]. This general science education requirement is often met by introductory astronomy classes as these classes have proven to be a popular choice by non-STEM major students [6–11]. As such, introductory astronomy classes often represent the last formal chance to influence the science knowledge, literacy, and interest of many students in their degree.

Students' understanding of the basic characteristics of celestial bodies and phenomena such as the day-night cycle, seasons, and phases of the Moon has been shown to be fundamental for understanding other scientific concepts and physical phenomena, and for making sense of modern research findings [1,4,5]. However, previous studies have repeatedly shown that students come to an astronomy course with a range of alternate conceptions about the way their reality works and that these misconceptions are strongly held and difficult to correct [4,12–15]. Indeed, research into student conceptual understanding of astronomy [4,13,14] has shown that a significant fraction of astronomy students retain their misconceptions, and are still unable to predict or explain basic astronomical observations even after a semester-long course. Students are not engaging with the typical traditional teaching techniques used in many astronomy courses of watching lectures about astronomy [12,16,17] and having certain features and phenomena pointed out to them on pre-arranged viewing nights. Students are capable of reproducing what they have been told, but they have not internalized the information and understood the concepts, and they cannot apply what they have learned to new situations.

The University of Western Australia runs two first year level introductory astronomy and space science units called *Our Universe* and *Our Solar System*. These units are open to STEM majors as an optional part of their course, and to non-STEM majors as "broadening" units to meet their general education degree requirements. The material in these units is split into two broad categories with *Our Universe* focusing on cosmology and astrophysics while *Our Solar System* focuses on astronomy and planetary science. The authors' experience teaching and running these units is consistent with the research that shows that a significant fraction of students retain their misconceptions and are still unable to correctly explain basic astronomical observations by the end of the course.

[*]david.gozzard@uwa.edu.au

Observational journals, requiring students to keep a diary of their own astronomical observations, have been proposed as a way to increase student engagement with course material and, as a result, improve students' conceptual understanding of astronomy and space science concepts [18]. Observation diaries and analogous activities have also been widely studied in other disciplines, and have been found to effectively engage students and lead to better learning outcomes, particularly if the work is assessed and the students are given marking rubrics to guide their efforts [19–21].

In this paper, we present an analysis of the contribution of these "observing diaries" to students' learning of astronomy concepts, and appreciation of astronomy's place in society and science in general. Students taking *Our Solar System* were required to keep a diary of their own astronomical observations and analyses over the duration of the semester. It was proposed that the diary would encourage students to undertake self-directed learning, since they must plan their observing schedule, draw or take images and photos, record data, and perform their own analyses and would not be able to take the information straight from lectures or online resources. The format of the diary was left up to the student to decide, allowing them to take ownership of their learning, express their creativity, and so increase their motivation for completing this task. The diaries were assessed at the end of the semester as a way of ensuring that the students completed the task, but also as a way of gauging the level to which the students engaged with, and understood, the astronomy principles that they were required to learn.

The aim of this research project was to find out whether requiring astronomy students to undertake more self-directed learning by keeping a regular diary of their own observations of the night sky increases their motivation and engagement with the course material and, as a result, improves their conceptual understanding of the basic principles of astronomy. To this end, an analysis of the students' attitudes towards astronomy and science was carried out, as well as an analysis of their conceptual understanding.

II. METHODS

A. *Our Solar System* and *Our Universe* course implementation

Students of the *Our Solar System* unit were set the task of keeping an astronomical observing diary which would be assessed at the end of semester. To provide a comparison group to assess the influence of the diary, students of *Our Universe* the following semester of the following year were not required to keep an observing diary but were asked to complete the same surveys and tests as the *Our Solar System* students.

Ideally, a better comparison would have been to have taught the *Our Solar System* units with and without the observing diaries. Unfortunately, due to changes in staffing and priorities, we were unable to do this, and so used the *Our Universe* unit as the next best comparator.

Instruction in both units was carried out in an identical manner. Each unit consisted of 24 lectures delivered in a traditional lecture manner; 12 noncompulsory tutorials where the tutor would cover a combination of set revision questions, assessment help, and student-chosen areas of confusion; and two field trips to the Gingin Gravity Discovery Centre (a public interactive science center and observatory). As is typical of introductory astronomy units open to non-STEM majors, the units adopt a largely concept-based approach to teaching with minimal mathematics. While *Our Solar System* focuses on astronomy and planetary science and *Our Universe* focuses on cosmology and astrophysics, as can be expected from these topics, there is significant overlap in the concepts that students were taught and expected to master in order to understand the course content. For example, a detailed understanding of the orbit of the Moon around the Earth and the planets around the Sun is required in both courses, for understanding basic astronomy and exoplanet detection in *Our Solar System*, and for understanding the development of the modern model of gravity and scientific view of the universe in *Our Universe*. Assessment of students' understanding of these basic astronomy concepts, as well as their attitudes towards science and astronomy, were used to determine the effectiveness of the observing diaries as a pedagogical tool.

There were $N = 137$ students in both *Our Solar System* and *Our Universe*. As is typical for an introductory astronomy unit open to non-STEM majors [22], the fraction of non-STEM majors (including arts, law, and business) in *Our Solar System* was greater than the fraction of STEM majors, with non-STEM majors making up 80% of the cohort. Engineering students made up the majority of the STEM majors with the remainder being a mix of other sciences (physics, chemistry, biology, psychology, etc.), mathematics, and pre-med students.

Unusually for an introductory astronomy unit open to non-STEM majors, the fraction of non-STEM majors in *Our Universe* was smaller than the fraction of STEM majors. Only 36% of students in *Our Universe* were non-STEM majors. Again, engineering students made up the majority of the STEM majors with the remainder being a mix of other sciences, mathematics, and premed students.

Apart from the assessed observing diary in *Our Solar System*, the two units had the same assessment format including written reports, an oral presentation, and a final exam comprising multiple choice and short-answer questions. The portion of total course marks allocated for each assessment were:

- Two 'field trip' reports each worth 15%. A total of 30%,
- A chosen topic report worth 20%,
- An oral presentation on the chosen topic worth 10%,
- The observing diary worth 5%, and
- The final unit exam worth 35%.

In *Our Universe*, the 5% of marks allocated to the observing diary in *Our Solar System* was re-allocated to the final exam. Both courses had a compulsory submission requirement, meaning a student who did not submit all of the assignments was assigned an ungraded fail for the course.

B. Observing diaries

All students in *Our Solar System* were required to keep an astronomy diary of their own, with regular astronomical observations throughout the semester, a total of 13 weeks of observing time between being assigned the task at the start of August and having to submit their diary for grading at the start of November. The diaries, worth 5% of their overall grade, were assessed at the end of semester and were graded on organization, effort, detail, depth of discussion and analysis, and initiative (going beyond the base requirements). The marking rubric (reproduced in Fig. 1) against which the students' work was assessed was provided to the students at the start of the semester. The observations were primarily intended to be done with the naked eye, along with optional simple instruments that students could build themselves, such as a quadrant built from a protractor and plumb bob, in order to ensure that all students had equitable access to resources. However, students were informed that they could use binoculars or a telescope if they owned one. The exact format of the diary (e.g., hard-copy journal, online blog, etc.) was left up to the student with the aim of increasing student engagement with the exercise by allowing them to exercise a good degree of their own creativity in the execution of the diary.

The students were shown exemplars of satisfactory and high scoring or otherwise impressive diaries from the previous year, some of which they also had access to online throughout the semester, and were advised to ask the tutor for guidance. The exemplars and the expectations and requirements of the diary assessment were discussed with the students in the first week of lectures and the students were informed that they could discuss their diaries with the course tutor (DRG) at any time. The tutor sought regular updates on diary progress and questions from students who attended the tutorials (tutorial attendance was not compulsory), and also provided students with online notification of up-coming astronomical events such as an eclipse or meteor shower. An astronomy night at a public observatory held early in the semester was also used as an opportunity to provide the students with guidance on how to make naked-eye astronomical observations. During tutorials, the students were shown the desktop astronomy app *Stellarium*, and were provided with a list of recommended websites and desktop and phone apps that would help them with their observations and identifying celestial objects.

The authors' experience with prescribing the observing diary assessment in previous years showed that keeping a journal is a novel experience for many students and students can experience confusion about how to execute their work, and anxiety about the assessor's expectations. To assist students to get started with making observations and gain familiarity with the night sky, the students were prescribed a set of compulsory observations to be completed in the first four weeks of semester. These compulsory observations were based on those suggested by Sadler and colleagues [18] and include:

Assessment scale: 5 — excellent, 4 — very good, 3 — satisfactory, 2 — acceptable, 1 — unacceptable

Diary clearly organized, i.e. used a consistent or systematic method of reporting observations	5	4	3	2	1	Diary disorganized
Interesting diary, e.g. contained images, photos, drawings, alt/az bearings	5	4	3	2	1	Boring diary. Did not contain:
Considerable effort, i.e. many (more than 10) detailed observations spread over semester	5	4	3	2	1	Very little effort, rushed with few observations.
Included reflections, comments, analysis or conclusions.	5	4	3	2	1	No analysis – just data.
Use of initiative/innovations Included compulsory observations	5	4	3	2	1	Unimaginative or little initiative, missing obs
TOTAL (/25) (scale to /5)						

FIG. 1. Marking rubric used to grade students' observing diaries.

- Observation of the position and time of four sunrises or four sunsets. (Nominally one observation per week.)
- Four observations of a planet moving against background stars. (Nominally one observation per week.)
- Four observations of the Moon on consecutive days/nights. (Observations do not have to be on consecutive days if prevented by adverse weather, but students should attempt to make their Moon observations spread across as few days as possible.)

A diary-focused discussion was held in the tutorial following the end of the compulsory observation period to assist students with the development of their observing practices and data analysis. However, the diaries were not formally assessed at this point.

C. Astronomy diagnostic test

The change in the students' conceptual understanding of basic astronomy principles was assessed using an *Astronomy Diagnostic Test* (ADT). Also known as a "misconceptions quiz," ADTs have been studied extensively as ways to gauge how well a course has confronted the misconceptions that university students bring to a class [15,23–30] as well as school children [31]. ADTs present test subjects with a series of multiple-choice questions designed to test their understanding of the target concepts, rather than their ability to reproduce rote-learned facts.

The ADT used in this study [32] was developed at the University of Sydney by O'Byrne and colleagues [30] for use in the southern hemisphere and was slightly modified by the authors for use in Perth, Western Australia. The ADT comprises 21 astronomy questions along with some further questions to collect demographic information. The students of *Our Solar System* and *Our Universe* were given the concept test online to complete before they had started their diaries, and again during the last week of the semester, after submitting their diaries. Comparing the precourse and postcourse ADT results between the two units allows a quantitative assessment of the effects of the observing diary to be made. A qualitative assessment of the change in students' conceptual understanding was also provided by reviewing the students' entries in their observing diaries. To encourage students to complete these online tests, a bonus of 2% added to their overall course grade was provided as an incentive for students who completed both ADTs.

The precourse and postcourse results for the ADT can be compared using a *normalized gain index*, $\langle g \rangle$ [33]:

$$\langle g \rangle = \frac{\text{avg post test result} - \text{avg pre test result}}{\text{maximum score} - \text{avg pre test result.}} \qquad (1)$$

This takes the ratio of the students' average achieved gain to the maximum possible average gain. A $\langle g \rangle = 0$ means no improvement while $\langle g \rangle = 1$ means the maximum possible improvement has been gained.

D. Science and astronomy attitude survey

As well as the diaries' effect on students' conceptual understanding, we also wanted to assess what effect the diaries had on students' appreciation of astronomy and its status as a science. To gauge students' attitudes towards astronomy and how those attitudes changed over the course of the semester, students in *Our Solar System* and *Our Universe* were also asked to complete an *Attitudes Towards Astronomy* survey at the beginning and end of the semester. The attitudes survey used for this program was developed by Zeilik and colleagues [33,34] and also includes questions to gauge the students' attitudes towards science in general, as distinct from astronomy. The attitudes survey records students' responses to a statement on a 5-point Likert scale, from strongly disagree to strongly agree. Half of the questions on the list are reversed. That is, a response of "5—strongly agree" indicates a positive attitude for half of the questions and a negative attitude for the other half. This encourages students to give considered responses, and somewhat negates the statistical effects of students who select the same option for every question in order to complete the questionnaire as quickly as possible.

Examples of the statements students were asked to respond to include:

- Astronomy is irrelevant to my life.
- Scientific skills make me more employable.

The 34 questions of this attitudes survey were added to the end of the ADT and students completed both questionnaires in the same online portal. An improvement in the students' attitude towards science and astronomy at the end of the semester would indicate that the students had engaged deeply with the course material. The observing diaries also provide a source from which to qualitatively gauge the level of student enthusiasm and engagement with the course material.

III. RESULTS

Tables I and II summarize the results of both the attitudes survey and the astronomy diagnostic test for *Our Universe* (without the observing diary) and *Our Solar System* (with the observing diary), respectively, while the results of the attitudes survey are presented in Fig. 2. The gain indices, $\langle g \rangle$, are calculated for the ADT as described in Eq. (1) and the effect size for each metric, the difference between the means of the postcourse and precourse distributions normalized by pooled standard deviation, are also given.

Students who had completed the other introductory astronomy unit prior to taking the one in which they were surveyed are excluded from the present analysis. Although there were 137 students enrolled in *Our Universe*, only 67 students completed both the precourse and postcourse questionnaires. For *Our Solar System*, which also had 137 students enrolled, only 117 valid observing diaries

TABLE I. Results of the pre- and postcourse surveys for the semester the diaries were not used.

Survey results without diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	48.4	17.8	14.3 90.5	54.4	19.1	14.3 100	0.33	0.12
Astronomy positivity index	1.4	1.0	−1.4 3.7	1.4	1.3	−2.4 4.2	0.07	
Science positivity index	2.1	1.2	−0.5 4.5	2.2	1.3	0.0 5.0	0.05	
Astronomy-science difference	1.0	0.8	−2.4 3.2	1.0	0.8	−2.2 2.8	−0.01	

TABLE II. Results of the pre- and postcourse surveys for the semester the diaries were in use.

Survey results with diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	49.2	18.7	9.5 95.2	58.8	18.3	23.8 100	0.52	0.19
Astronomy positivity index	1.4	0.9	−0.8 3.2	1.6	1.2	−2.4 3.7	0.20	
Science positivity index	2.5	1.3	−1.6 4.8	2.5	1.2	−0.2 4.5	0.04	
Astronomy-science difference	1.2	0.9	−3.0 4.0	1.1	0.8	−2.0 3.2	0.12	

were submitted for grading and only 71 students completed both the precourse and postcourse questionnaires.

The results presented below are discussed in Sec. IV.

A. Astronomy diagnostic test

At the start of the semester, the average score for the ADT for students enrolled in *Our Universe* (without diaries as an assessment) was 48.4% with a standard deviation of 17.8. Student scores ranged from 90.5% to 14.3%. The average score for the ADT for students enrolled in *Our Solar System* (with diaries as an assessment) was 49.2% with a standard deviation of 18.7. Student scores ranged from 95.2% to 9.5%. These results are very similar and provide a good basis on which to make a comparison of the students' improvement with and without the observing diaries.

At the end of the semester the average score for the ADT without diaries was 54.4% with a standard deviation of 19.1. Student scores ranged from 100% to 14.3%. This corresponds to a gain index of $\langle g \rangle = 0.12$ with an effect size of 0.33. The average score for the ADT with diaries was 58.8% with a standard deviation of 18.3. Student scores ranged from 100% to 23.8%. This gives a gain index of $\langle g \rangle = 0.19$ with a significant effect size of 0.52.

Between the precourse and postcourse ADT results for the units without and with the observing diary, we find an improvement in ADT gain index from 0.12 to 0.19, and significant increase in effect size from 0.33 to 0.52. These results are summarized in Tables I and II.

These results can be broken down further to analyze the difference between students studying a STEM major and students studying a non-STEM major. In *Our Universe* (without diaries) the pre- and post-course average ADT score for STEM students increased from 52% to 57%, a gain of only $\langle g \rangle = 0.10$ with an effect size of 0.29, while the average ADT score for the non-STEM students had a greater increase from 42% to 50.4%, a gain of $\langle g \rangle = 0.14$ with an effect size of 0.45.

In *Our Solar System* (with diaries), the average ADT for STEM students increased from 56.5% to 65.6%, a significantly greater gain of $\langle g \rangle = 0.21$ with an effect size of 0.64, while the average ADT score for the non-STEM students had increased from 47.5% to 57.3%, a gain of $\langle g \rangle = 0.19$ with an effect size of 0.52.

Results for STEM major students are summarized in Tables III and IV while results for non-STEM major students are summarized in Tables V and VI. Based on these ADT results, the inclusion of the observing diary in the course work might have led to a moderate increase in students' understanding of the desired basic astronomy concepts.

B. Survey of attitudes towards astronomy

1. Whole cohort

The results of the attitude survey were separated into "astronomy" attitudes and "science" attitudes and compiled into a *positivity index* ranging from −5 to 5, with −5 indicating strongly negative attitudes towards the subject and 5 indicating strongly positive attitudes towards the subject. The results from the precourse and postcourse attitude surveys for the units both with and without the observing diaries are shown in Fig. 2, and are summarized in Tables I and II.

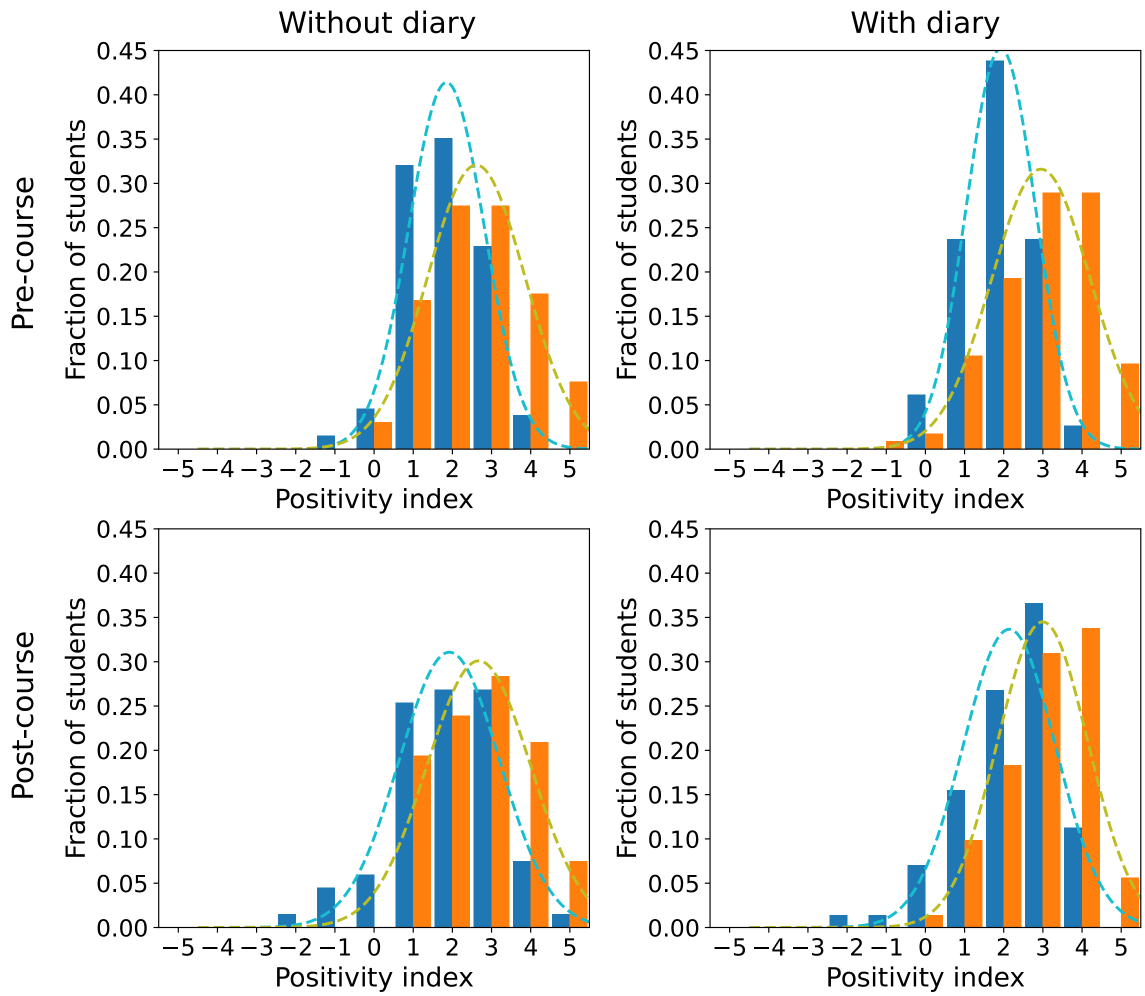

FIG. 2. Histogram of students' attitudes towards astronomy (blue) and science (orange) precourse (top) and postcourse (bottom) for *Our Universe* (without observing diary, left) and *Our Solar System* (with observing diary, right) from the results of the attitudes towards astronomy survey. The dashed curves (astronomy, cyan; science, olive) show a normal distribution (bell curve) fitted to the corresponding histogram. (Note that the bell curves have been shifted right by 0.5 to remove the "binning" effect and align visually with the underlying histogram.)

Students' attitudes towards both astronomy and science in the precourse survey were overall positive, as would be expected for a cohort of students who had elected to take these optional units. However, a high positivity index in one field does not necessarily correspond to a high positivity index in the other, with some students having very positive feelings about astronomy, but neutral feelings about science, or vice versa. The average separation between students' positivity index values for astronomy versus science was 1.0 index point for *Our Universe* and 1.2 index points for *Our Solar System*. This indicates that for a large fraction of students there was a significant

TABLE III. Results of the pre- and postcourse surveys for the semester the diaries were not used. (STEM major students).

Survey results without diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	52.0	17.3	14.3 90.5	57.0	17.3	19 90.5	0.29	0.10
Astronomy positivity index	1.5	0.9	−1.4 3.7	1.7	1.4	−2.4 4.2	0.18	
Science positivity index	2.5	1.0	−0.5 4.5	2.7	1.3	0.0 5.0	0.09	
Astronomy-science difference	1.2	0.9	−2.4 3.2	1.1	0.8	−1.2 2.8	0.12	

TABLE IV. Results of the pre- and postcourse surveys for the semester the diaries were in use. (STEM major students).

Survey results with diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	56.5	14.2	33.3 90.5	65.6	14.4	28.6 85.7	0.64	0.21
Astronomy positivity index	1.6	0.6	0.4 2.7	1.8	1.2	−2.4 3.2	0.21	
Science positivity index	3.2	0.7	1.4 4.5	2.7	1.0	0.0 4.5	−0.60	
Astronomy-science difference	1.6	0.9	−0.5 3.6	1.1	0.8	−0.7 2.8	0.60	

TABLE V. Results of the pre- and postcourse surveys for the semester the diaries were not used. (non-STEM major students).

Survey results without diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	42.0	16.5	14.3 85.7	50.4	20.6	14.3 100	0.46	0.14
Astronomy positivity index	1.0	0.9	−1.2 3.5	1.0	0.9	−1.2 2.7	0.00	
Science positivity index	1.2	1.0	−0.5 3.9	1.4	1.0	0.0 3.6	0.20	
Astronomy-science difference	0.6	0.4	−1.8 1.4	0.9	0.7	−2.2 2.0	−0.56	

TABLE VI. Results of the pre- and postcourse surveys for the semester the diaries were in use. (non-STEM major students).

Survey results with diaries								
	Precourse			Postcourse				
Survey category	avg	st-dev	range	avg	st-dev	range	Effect size	Gain index
ADT score (%)	47.5	19.1	9.5 95.2	57.3	18.6	23.8 100	0.52	0.19
Astronomy positivity index	1.4	0.9	−0.8 3.2	1.6	1.2	−1.1 3.7	0.20	
Science positivity index	2.3	1.3	−1.6 4.7	2.4	1.2	−0.2 4.3	0.08	
Astronomy-science difference	1.1	0.8	−3.0 4.0	1.1	0.8	−2.0 3.2	0.00	

conceptual disconnect between science and astronomy. The average beginner student of *Our Universe* or *Our Solar System* seems to believe that astronomy is both more difficult and less applicable than their concept of “science.”

Analyzing the post-course tests for both units, the results show that for the unit with diaries, students’ attitudes towards astronomy received a significant increase in positivity (effect size $= 0.2$) while the effect was very small (effect size $= 0.07$) for the unit without diaries. The results also show that, for both units, the range of attitudes towards astronomy has increased, with some students giving significantly lower positivity scores than in their precourse survey. This is unfortunate, but it is inevitable that there will always be some students for whom a course does not meet their prior expectations, or for whom the concepts presented at a tertiary level are difficult to grasp, and this may be the cause of their disillusionment with astronomy.

2. STEM vs non-STEM majors

The results summarized in Fig. 2 and in Tables I and II can be further broken down into results for STEM majors and for non-STEM majors. Figure 3 shows the difference in attitudes of STEM majors (blue) and non-STEM majors (orange) towards astronomy, while Fig. 4 shows the difference in attitudes of STEM majors and non-STEM majors towards science, pre- and postcourse, with and without the observing diary.

The results from the precourse and postcourse astronomy diagnostic test and attitude surveys for students undertaking STEM majors in the units both with and without the observing diaries are summarized in Tables III and IV, while the same results for students undertaking non-STEM majors are summarized in Tables V and VI.

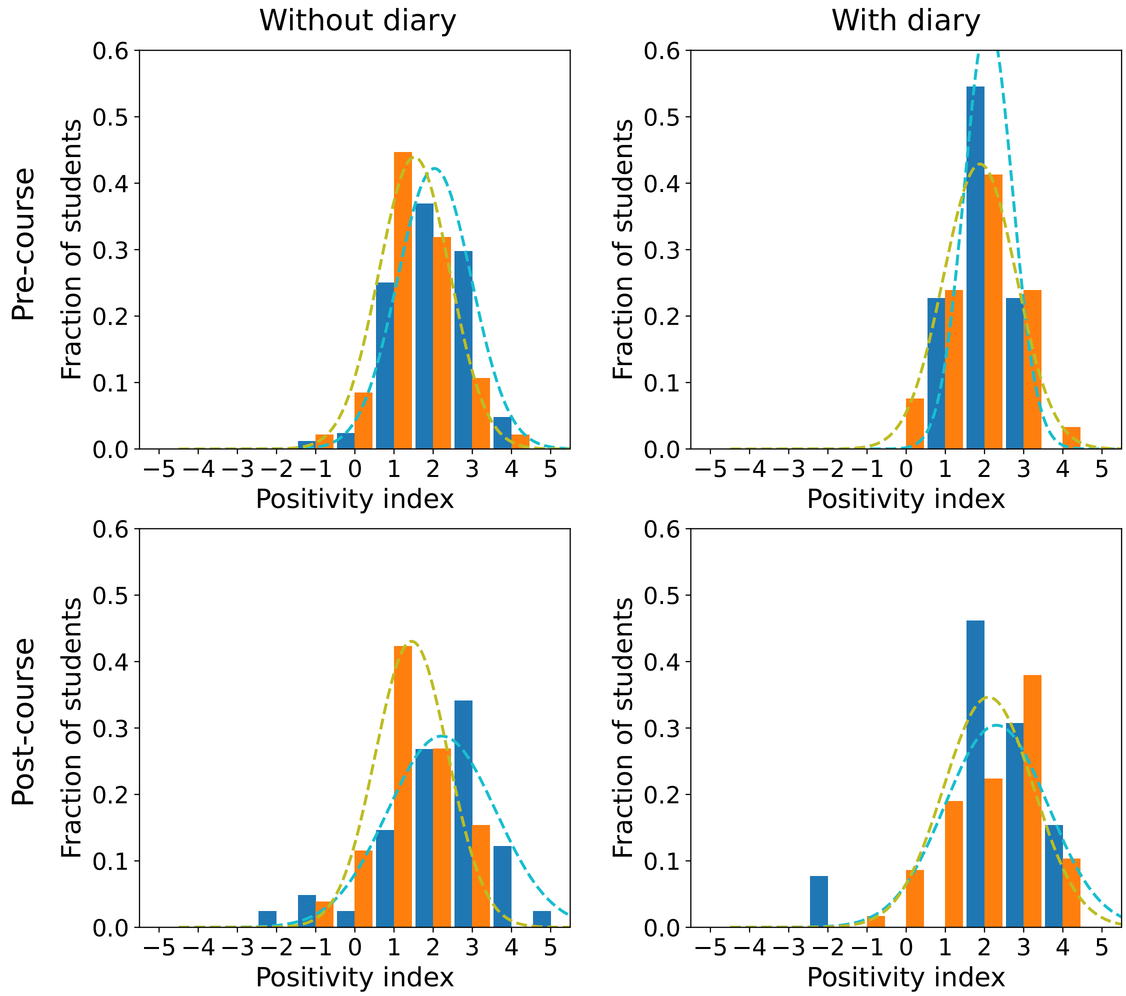

FIG. 3. Histogram shows STEM major (blue) versus non-STEM major (orange) students' attitudes towards astronomy precourse (top) and postcourse (bottom) for *Our Universe* (without observing diary, left) and *Our Solar System* (with observing diary, right) from the results of the attitudes towards astronomy survey. The dashed curves (STEM-majors, cyan; non-STEM majors, olive) show a normal distribution (bell curve) fitted to the corresponding histogram. (Note that the bell curves have been shifted right by 0.5 to remove the binning effect and align visually with the underlying histogram.)

C. Exam vs diary performance

Students who perform well in exam-style assessments do not always show equivalent prowess at more unstructured, open-ended, self-directed, or hands-on assessments [11,18]. To determine how the observing diary assessment compares to more traditional assessment methods, we can compare students' performance in the diary against their performance in the unit exam. Figure 5 shows a correlation plot of students' marks for the observing diary against their marks for the final unit exam at the end of the semester.

Students who did not submit an observing diary or scored zero for any reason (late submission, blank submission, etc.) have been removed from the plot, leaving the $N = 117$ students who submitted the diary. A line of best fit through the data has been plotted using the least-squares method.

The gradient of the trend line (fitted using the least-squares method) is 0.31, indicating that, in general, students achieved a higher mark in the diary than they did in the exam, and that students who did well in the diary tended to do well in the exam. However, there is a lot of scatter in the data and the R^2 correlation value is only 0.2, indicating that students who achieved high marks in the conventional exam did not necessarily achieve correspondingly high marks in the diary, reflecting the self-directed nature of the diary assessment.

D. Student responses to the observing diaries

A qualitative analysis of the submitted observing diaries themselves also offers insights into the dairies' effectiveness as an educational tool. The majority of students rose to the challenge and submitted a high standard of work. The average mark was 82% and indicates the level of effort that students put into the diary, despite the diary only being worth 5% of their total grade, which suggests that the students found it an engaging learning experience. The

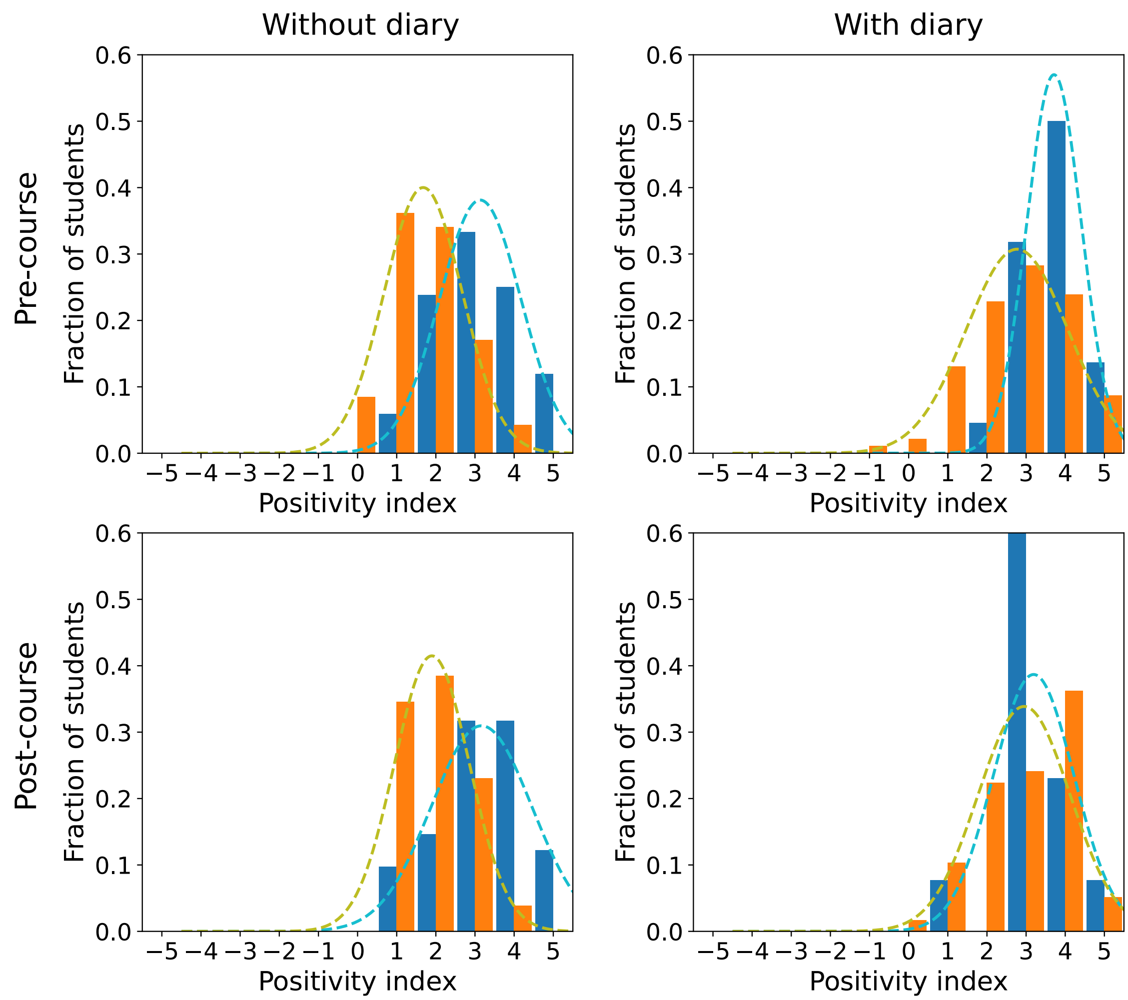

FIG. 4. Histogram shows STEM major (blue) versus non-STEM major (orange) students' attitudes towards science precourse (top) and postcourse (bottom) for *Our Universe* (without observing diary, left) and *Our Solar System* (with observing diary, right) from the results of the attitudes towards astronomy survey. The dashed curves (STEM-majors, cyan; non-STEM majors, olive) show a normal distribution (bell curve) fitted to the corresponding histogram. (Note that the bell curves have been shifted right by 0.5 to remove the binning effect and align visually with the underlying histogram.)

marks ranged from 36% to 100% and 22 of the 117 diaries were awarded 100% against the marking rubric shown in Fig. 1.

Seven of the students stated in their diary how much they enjoyed working on the diary, with a further four stating their enjoyment in emails to the tutor and course co-ordinator. Three of the comments were

- I really enjoyed this assignment because it got me outside and looking at the sky, which was really relaxing.
- I didn't think I would enjoy the observing assessment, but I did.
- I never noticed how much the sky changes. It's amazing!

However, two students used the end-of-semester anonymous course feedback opportunity to express their dislike for the diary assignment, stating that the need to make the observations was inconvenient and that they failed to see

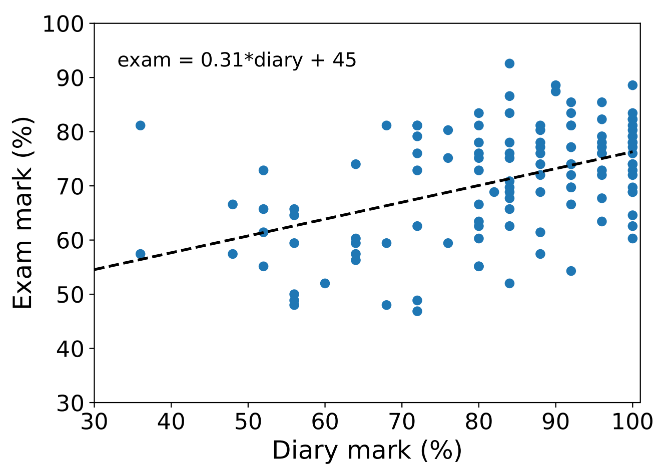

FIG. 5. Performance in the observing diary assessment versus the final course exam for each student who completed both assessments, $N = 117$. The dashed black trend line shows a least-squares fit to the data.

how the diary was a useful or fruitful learning endeavor. While several students left positive comments about the unit as a whole and complimented their favorite lecturers, no students left positive comments about the observing diary task.

As had been planned, the diaries gave student the opportunity to express their own preferences and creativity. Sixty-two of the 117 diaries were electronically submitted while the rest were handwritten or other hard copy. The majority of diaries either started with or quickly developed a simple but sound observing template that the student would fill out each observing session. Several of the diaries were very artistic, having either detailed drawings of observations or decorations such as intricate drawings of characters associated with the constellations. Three of the diaries were in the form of online blogs that were publicly accessible. (Two of these students indicated a desire to continue blogging their observations after the end of semester. However, only one did, and only for a few weeks.) Two students compiled their observations into a Microsoft PowerPoint presentation (one of which used timings to automatically play itself), using the animation features of PowerPoint to highlight various observations and analysis.

More broadly, a few distinct types and styles of observing diary emerged. Table VII shows a breakdown of these different diary types. The most prominent type was a "journal" in which students recorded separate entries for each of their observing sessions. The prevalence of this style of diary is unsurprising because it is, perhaps, the most intuitive way to format an observing diary, and the example diaries that the students were shown predominantly had this format. The journal type included three subtypes: handwritten with hand-drawn figures (55), typed journals with inserted digital images (31), and hybrid journals with typed text and hand-drawn figures (5). The aforementioned artistic or illustrated (5), blog (3) and PowerPoint (multimedia, 2) diaries also exhibited the journal format.

The "table" type diary (18) is related to the journal type, but only presented tables of recorded data for each observation, with little or no written information or discussion. These diaries were scored very poorly against the discussion or analysis and initiative or innovations marking criteria from Fig. 1.

TABLE VII. Breakdown of types of observing diary submitted. ($N = 117$ diaries).

Diary type or style	Number of diaries
Handwritten journal	55
Digital journal	31
Hybrid journal	5
Table	18
Report	3
Artistic or illustrated	5
Multimedia	2
Blog	3
Over-reliance on software	19

Three of the digitally formatted submissions did not follow the (expected) diary or journal format. Instead, these students wrote up their observations and analysis as a report, very similar in presentation to a first-year undergraduate laboratory report.

The different types of diaries were given equal weight when graded as long as they were well organized and had a clear and systematic method of recording observations. (The first marking criterion in Fig. 1.) However, the artistic diaries, blogs, and PowerPoint diaries were rewarded for their initiative and innovation. (The last marking criterion in Fig. 1.)

Nineteen (19) of the diaries show that the student had an over-reliance on astronomy software to make their observations. (In a few cases this constituted falsification of results and is discussed further in Sec. IV. B below.) Features of diaries in this category include

- A majority of the reported observations being taken directly from astronomy apps, rather than being the student's own viewing of the night sky. These diaries often cited cloudy weather as a reason for taking data from apps. Students had been told that, if weather prevented them from making a successful observation, they were to try repeating the observation on another night during the week, but that they were permitted to fill in the occasional missing data point if several days of cloudy weather (or other considerations such as illness) prevented them from making a successful observation themselves.
- All position data of celestial objects being taken from astronomy apps, with no attempt to make measurements from the student's point of view, such as angle of the object from a landmark.
- Confusion if the student's own view of the sky differed from that shown in the astronomy app. Often the student would then attribute this difference to their own observations of the sky being in error, instead of trying to work out the reason for the difference, and they would rely on writing down what the app indicated they should be seeing.
- Use of built-in features of an astronomy app to "plot" star trails or the path of a planet instead of the student performing their own analysis of their data.

These students' overreliance on astronomy apps to show them what they "should be seeing" was usually detrimental to the students' understanding of the phenomena they were observing, possibly because they were not effectively engaging with the project and seriously considering the data, instead just relying on an app to show them the "right answer." These diaries were usually rewarded for having a well-organized and systematic method of recording observations, but were penalized for lack of proper discussion or analysis and initiative or innovation.

TABLE VIII. Breakdown of diary mark per marking rubric criterion. ($N = 117$ diaries).

	Organized diary	Interesting diary	Effort	Analysis and discussion	Initiative or innovation
Mean	4.8	4.4	4.0	3.4	3.9
Median	5	5	5	3	4
Mode	5	5	5	2	5
Minimum	1	1	1	1	1
Maximum	5	5	5	5	5

The observations and details presented in the diaries show that, for the most part, students noticed the astronomical cycles and phenomena that were the goal of setting the observing diary as an assessment. However, the majority of students failed to follow through with analysis and conclusions based on their observations. Table VIII shows the breakdown of diary marks per marking rubric item.

The 22 students who were awarded 100% presented clear and detailed weekly observations, and followed up with an analysis and discussion of their data towards the end of semester. Some students performed some basic analyses of their data, but failed to discuss their analyses or show that they had put any effort into trying to interpret what their data was showing them. For example, three students made basic altitude-azimuth observations of a few bright stars at the same time each night over the course of the semester and then plotted their recorded values. They noted that most of the stars traced an arc in their plots, but that one star traced an almost straight line. Instead of discussing this feature and eventually realizing that this particular star was much closer to the equator than the others, all of these students put this apparent anomaly down to errors in their own observations.

The majority of students made a significant number of thorough observations, but made little or no effort to discuss or analyze their observations. However, these students still received marks around 70% to 80% because the criteria in the marking rubric made it possible to achieve this mark by doing good observations but with little effort invested in considering the observations in greater depth.

As stated previously in Sec. II. B, a tutorial class discussing the diaries, students' progress, and data analysis was held following the end of the compulsory observation period. Students were invited to show their diaries to other students in the tutorial group and discuss their findings. This class was used as an opportunity for the tutor to assist the students to improve their observing practices and guide the students in how to analyze their data. However, following this tutorial, very few students changed their observing habits after receiving feedback.

IV. DISCUSSION

A. Observing diaries as a positive teaching strategy

Comparing the results of the precourse and postcourse ADT for *Our Universe* and *Our Solar System* allows us to quantitatively assess the impact the observing diary assignment had on students' learning of basic astronomy concepts. Taking the student cohort as a whole, the gain index for the unit with the diaries was significantly greater than for the unit without the diaries (0.19 with the diaries compared to 0.12 without). However, the gain indices for both units are very low compared with other interactive engagement courses, and even the unit with the diaries only achieved a gain index close to that of courses taught using traditional teaching methods [33]. This is probably due to the low weighting of the mark for the diary (5%) and may reflect an emphasis on facts, rather than understanding of concepts, in the unit as a whole, plus the unit was taught in a largely traditional mode. In spite of this, the significant increase in gain index for the unit with the diary assessment suggests that the diaries have a positive effect on students' learning of basic astronomy concepts.

Considering the ADT results for STEM major and non-STEM major students separately shows some differences between the two groups of students. As might be expected, STEM majors showed a greater understanding of astronomy concepts than non-STEM majors both at the start and end of the course. However, for STEM majors, the gain in ADT improvement more than doubled from 0.10 without dairies to 0.21 with diaries, while the gain for non-STEM majors was less significant, increasing from 0.14 without diaries to 0.19 with diaries. For the unit without the observing diary, non-STEM majors gain more than STEM majors, while for the unit with the diary, STEM and non-STEM majors gain a similar amount (0.21 versus 0.19 respectively). This suggests that STEM majors might be getting a significant benefit in conceptual understanding due to the diary project, while non-STEM majors are deriving greater benefit from the unit as a whole. This fits with the fact that the units are first-year introductory astronomy units and STEM students are likely to come to the unit with greater background knowledge, which the non-STEM students must catch up on via the course material.

The attitudes towards astronomy survey shows that the average student of *Our Universe* and *Our Solar System* enters the unit believing that there is a significant conceptual disconnect of astronomy from "science" and that astronomy is both more difficult to learn and less applicable to real-world problems than science. At the end of *Our Universe* (without the diary), these attitudes had changed very little for STEM students. Non-STEM students registered negligible change in their attitude towards astronomy, but a small improvement in their attitude towards science, leading to a significant increase in their dissociation between science and "astronomy."

However, for *Our Solar System* (with the diaries) we see a small but significant improvement in students' attitudes towards astronomy, resulting in smaller difference between their appreciations of astronomy and science. This change is driven mostly by STEM major students who, on average, registered a small increase in positive attitudes towards astronomy, but a significant decrease in positive attitudes towards science in general. This is likely to be affected by the small sample size where only 13 STEM-major students completed both the pre- and postcourse survey, two of whom reported significant drops in their positive attitudes towards both astronomy, and science in general. This highlights a limitation of the analysis presented here. Unlike the non-STEM majors, students studying STEM degrees will have been taking other STEM subjects during the semester and their experiences in these other classes will have impacted their positive or negative feelings towards science and astronomy.

The results of the attitudes survey show that the inclusion of the observing diary in the course might have had a positive effect on students' attitudes towards astronomy as a science, and astronomy in general. While this is effect is not large, it provides a basis and direction for improvement of students' engagement with the unit as a whole.

This conclusion is also reflected in the qualitative analysis of students' responses to the observing diary assignment. A significant subset of students put a lot of effort into their observations, record keeping, and analysis and, as a result, had a very positive experience in the diary and the unit as a whole. A smaller subset of students put little effort into the diary and could not see the point of the exercise, while the majority of students performed very good observations and record keeping, but failed to invest the time and effort required to analyze their data and think more deeply about what they were seeing. In any similar future study, follow-up interviews with students who underperform in the data analysis might shed light as to why this is the case. We surmise that this result is due to the tendency of students to prefer to do work that is "fun" over work that is harder, choosing to put a lot of effort into the taking and layout of observations, but losing interest when it came to the more difficult task of analyzing and discussing their data. The high average mark is indicative of the effort that the majority of students put into the diary, but this clearly needs to be followed up with greater encouragement and explanation of how to perform analysis of the data and really engage with the concepts.

The self-directed nature of the observing diary meant that, as a learning opportunity, it was very much a case of "you reap what you sow." The more effort a student put into their observations, background reading, and analysis, the more they got out of the diary in terms of new insights, understanding, and knowledge. This was very apparent in the marks, with scores strongly reflecting the amount of effort the students put in to the diary.

The compulsory observations and corresponding tutorial showed that students were very resistant to changing their observing habits, documentation, or analysis plans after receiving feedback from the tutor, indicating that students had already become very comfortable with their process by this point and were reluctant to change it. For future teaching improvement it will be necessary to provide students with greater feedback or guidance at a much earlier stage, or to provide a greater incentive for changing their observing process, such as providing a mid-term or "predicted" mark for their diary following the end of the compulsory observations.

Comparing students' marks for the observing diary against their performance in the final unit exam shows a weak correlation between performance in one assessment versus the other. This suggests that the self-directed nature of the observing diary assessment requires a very different type and level of effort from students compared with more traditional assignments. The diaries may be a very positive experience for students who do not suit conventional study and exam assessment techniques. However, the diaries seem to have exposed weaknesses in students who are more comfortable with traditional practices. It may be argued that this is a good thing, given that the unit with the diaries showed a greater improvement in students' understanding of the desired astronomy concepts than the unit without the diaries.

Overall, the results suggest that the observing diary had a positive impact on students' conceptual learning and was a positive tool for student learning and engagement that contributed to students' more positive view of astronomy overall and as a science.

A number of variables contribute to the uncertainty in this conclusion, both in terms of how these variables have influenced the students' responses, and their impact on practical aspects of the analysis. As noted previously, the number of STEM major students who completed both the pre- and postcourse surveys for *Our Solar System* was only 13, leading to weak statistical significance of the resulting data. While the analysis presented here sought to test astronomy concepts that are common to both *Our Solar System* and *Our Universe*, and lectures on the tested concepts, such as Kepler's laws, were common to both units, remaining differences in the two units contribute to the uncertainty. These contributions are not just limited to the course content, but also include the difference in the cohorts (significantly more STEM students in *Our Universe*), student's attitudes towards different lecturers in the two units, and differences in when certain concepts are presented during the semester. Student's enthusiasm for the observing diary project and resulting time-on-task will have influenced the outcome. Additionally, possibly particularly for the STEM students, students' experience in their other units during the semester may have significantly affected their attitudes towards science and astronomy. The ADT, including the questions designed to obtain demographic information, plus the attitude survey made the precourse and postcourse questionnaires nearly 60 questions long. Especially at the end of the semester, when students are stressed, studying for exams, and being given various quality of teaching evaluations and questionnaires in other classes, they may be experiencing a large degree of survey fatigue, which may have impacted the results of the postcourse ADT and attitude survey. This is likely to be the reason many students chose not fill out the postcourse survey. Since performing the investigation described in this paper, the authors have moved on to new roles and no longer have access to *Our Solar System* and *Our Universe* units to perform further research.

While in this paper we have focused on the inclusion of students' observation diaries in the context of astronomy education, research has been done on analogous activities in other disciplines. For example, students who were required to write field notes and sketches demonstrated significant learning benefits [19,20]. A metastudy of a large number of articles showed that students who draw by hand using pen and pencil have significant benefits in terms of their learning [21]. Teaching students the scientific practice of effective observation is regarded as a crucial skill in disciplines such as biology [35], and this is also true of astronomy. These and other research (for example, Refs. [36,37] and the references therein) reinforce the well-known strategy that students undertaking hands-on activities which clearly engage students does lead to better learning outcomes, particularly if the work is assessed and the students are given clear instructions and expectations such as rubrics, to guide their efforts. Our work is consistent with these conclusions.

B. Lessons learned

Experience gained from teaching the unit with the diaries and this exploratory analysis of their impact on student engagement and learning has resulted in a number of "lessons learned" and suggested future improvements to the observing diary assessment.

Mark compulsory observations early.—Following the compulsory observations, the diaries should be marked, or given a 'predicted mark', to give students a better idea of their progress with the assessment and greater encouragement to alter their practices and perform a deeper analysis of their observational data.

Changes to the marking rubric.—The marking rubric is too generous and does not sufficiently discriminate high quality diaries from more mediocre ones. The average mark was 82%. Students were able to achieve a mark of 70% without doing any data analysis, and the fact that students could achieve a mark of 20% for submitting "unacceptable" work increased the overall average significantly. The marking rubric should be altered to reduce the ease with which students can gain marks for doing little work against a particular criterion, and should increase the weighting of the analysis criterion in order to encourage students to put more effort into data analysis and discussion. Greater effort also needs to be made to communicate to the students that analysis of data and discussions of observations and findings is extremely important and students should not simply limit their work to recording observations. Creating

Assessment scale: 5 - excellent, 4 - very good, 3 - satisfactory, 2 - acceptable, 1 – unacceptable, **0 – not done**

Diary clearly organized, i.e. used a consistent or systematic method of reporting observations	5	4	3	2	1	**0**	Diary disorganized
Interesting diary, e.g. contained images, photos, drawings, alt/az bearings	5	4	3	2	1	**0**	Boring diary. Did not contain:
Considerable effort, i.e. many (more than 10) detailed observations spread over semester. Compulsory obs done.	5	4	3	2	1	**0**	Very little effort, rushed with few observations.
Included reflections/comments on observations and analysis of data.	5	4	3	2	1	**0**	**No comments or analysis – just data.**
Included discussion of the analysis and conclusion drawn.	5	4	3	2	1	**0**	**No discussion or conclusions**
TOTAL (/25)							

FIG. 6. Suggested improved marking rubric to grade observing diaries. Changes from the original rubric presented in Fig. 1 are shown in bold.

"learning moments" by asking students to relate their observations to concepts covered in class may assist with encouraging them to critically consider their observations. Figure 6 shows a suggested improved marking rubric, which reduces the marks that students are able to get for doing minimal or no work against a criterion, and increases the weighting of the data analysis and discussion by splitting analysis of data and discussion and conclusions from the analysis into two separate criteria. This will better distinguish students who attempt a more thorough analysis and discussion from students who perform a detailed analysis, such as plotting the progress of objects through the sky over the observing period, but who do not follow up with discussion of their results, hoping that the marker will draw the desired conclusions and will award them the marks anyway. The removal of the "initiative or innovations" criterion may also be beneficial because students often object to vague marking criteria, and a diary that is outstanding in its execution can be rewarded against the other criteria.

Increase diary marks weighting.—The observing diary was only worth 5% of students' total grade for the unit. The weighting of the diary should be increased to 20% to 30% to better reflect the amount of effort that many students put into their diary, and to encourage more students to take the assignment more seriously and invest more effort in the data analysis.

Guided observing night.—While the compulsory observations and observatory night were useful opportunities to guide students in what they should be observing and how they should be doing it, the fact that students' observing practices quickly become ingrained means that a teacher-lead observing session very early in the semester may be valuable in increasing students' confidence in their observing and analysis throughout the rest of the semester.

Draw diagrams.—Encouraging students to draw diagrams of what they think is going on in their observations (e.g., relative orbital positions of the Earth and observed object to the Sun) might facilitate deeper thought and understanding of the phenomena in question [12].

Fake observations using astronomy apps.—A small number of students used the recommended astronomy apps and websites to falsify some or all of their observations. These false observations were surprisingly easy to catch while assessing the diaries, generally because they highlighted the student's lack of understanding of the phenomena in the night sky. Some errors that gave students away include

- Describing observations that could only have been made from the northern hemisphere (*Stellarium's* default observing position is Paris) or from a geographic location that was clearly not the student's claimed observing location.
- Claiming to have seen extremely faint or deep sky objects such as comets or Pluto with the naked eye (and from light-polluted inner suburbia).
- Claiming to record the angular positions of stars and planets to better than hundredths of a degree using only a compass and DIY inclinometer.
- Claiming to have witnessed events that had not been possible to view from their stated location due to things such as adverse weather. (For example, a student claimed to have witnessed a lunar eclipse that had been completely obscured by thick cloud across the whole Perth region.)

While falsification of observations is disappointing, only a small number of students did this and instances of falsification were easy to catch and clearly showed the student in question was failing to understand basic astronomy concepts as a result. An interim assessment and marking of the diaries at an early stage in the semester, such as after the compulsory observations will provide an opportunity to detect false observations earlier and demonstrate to the student that they will be caught out and have marks deducted for the practice.

V. CONCLUSION

Students in a first-year university introductory astronomy unit were instructed to perform self-directed naked-eye astronomical observations and to record and analyze their observations in an observing diary over the course of the semester. An astronomy concepts diagnostic test and attitudes towards astronomy and science survey were administered to the students at the start and end of the semester to gauge the change in students' conceptual understanding of basic astronomy concepts, and their attitudes towards astronomy and science in general. To determine what effect the observing diary had on students' learning and appreciation of the subject, the same precourse and postcourse tests were administered to a second similar introductory astronomy unit that did not include the observing diary assessment.

The ADT showed greater gains for the class with the observing diaries, and the attitude survey shows that the class with the diaries also developed a greater appreciation of, and more positive attitude towards, astronomy compared with science in general. The level of effort and creativity that the majority of students put into their diary shows that most students rose to the challenge of this self-directed learning opportunity and had greater engagement with the course material and key concepts than the nondiary class. A comparison of students' marks for the diary versus their marks in the final unit exam showed little correlation between students' scores for the two assessments, indicating that this nontraditional form of assessment may benefit students who struggle with the traditional forms of study and assessment. Overall, the results suggest that the observing diary was a positive learning experience, having a positive impact on students' conceptual learning, as well as contributing to an improvement in students' perception of astronomy as a subject.

This work recommends the use of naked-eye astronomy observing diaries as an educational tool in introductory astronomy units. Suggested improvements to the execution of the diary assessment from an educator's point of view include increasing the fraction of unit marks the diary is worth to between 20% and 30%. Coupled with greater emphasis on the analysis and discussion of the recorded observations, this might encourage students to invest more effort in thinking deeply about and understanding the phenomena they are seeing. A guided observing session early in the semester may also help students to feel more confident about performing their observations and encourage them to further analyze their data.

ACKNOWLEDGMENTS

This work was supported by the University of Western Australia's Postgraduate Teaching Internship Scheme. The authors would like to thank Professor David Blair and Ms. Ruby Chan for their support and for enabling this investigation to be run in the *Our Solar System* and *Our Universe* units. D. R. G. would also like to acknowledge the support of the Forrest Research Foundation.

---

[1] National Research Council, *Astronomy and Astrophysics in the New Millennium* (The National Academies Press, Washington, DC, 2002).
[2] A. Hobson, The surprising effectiveness of college scientific literacy courses, Phys. Teach. **46**, 404 (2008).
[3] D. Wittman, Shaping attitudes toward science in an introductory astronomy course, Phys. Teach. **47**, 591 (2009).
[4] P. M. Sadler, H. Coyle, J. L. Miller, N. Cook-Smith, M. Dussault, and R. R. Gould, The Astronomy and Space Science Concept Inventory: Development and validation of assessment instruments aligned with the K-12 national science standards, Astron. Educ. Rev. **8**, 010111 (2010).
[5] M. Apartzaki and M. Kalogiannakis, Astronomy in early childhood education: A concept-based approach, Early Child. Educ. J. **44**, 169 (2016).
[6] J. Blue, Assessing astronomy students' views about the nature of scientific enquiry, Phys. Rev. Phys. Educ. Res. **14**, 010141 (2018).
[7] S. R. Buxner, C. R. Impey, J. Romine, and M. Nieberding, Linking introductory astronomy students' basic science knowledge, beliefs, attitudes, sources of information, and information literacy, Phys. Rev. Phys. Educ. Res. **14**, 010142 (2018).
[8] A. Fraknoi, The state of astronomy education in the US, in *Astronomy Education: Current Developments, Future Coordination*, edited by J. A. Percy (Astronomical Society of the Pacific, San Francisco, CA, 1996).
[9] B. Partridge and G. Greenstein, Goals for "Astro 101:" Report on workshops for department leaders, Astron. Educ. Rev. **2**, 46 (2004).
[10] H. Tignaneli and Y. Benétrau-Dupin, Perspectives of history and philosophy on teaching astronomy, in *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, Dordrecht, 2014), p. 603.
[11] M. M. Wooten, K. Coble, A. W. Puckett, and T. Rector, Investigating introductory astronomy students' perceived impacts from participation in course-based undergraduate research experiences, Phys. Rev. Phys. Educ. Res. **14**, 010151 (2018).
[12] B. Martinez Pena and M. J. Gil Quilez, The importance of images in astronomy education, Int. J. Sci. Educ. **23**, 1125 (2001).
[13] J. M. Bailey and T. F. Slater, A review of astronomy education research, Astron. Educ. Rev. **2**, 20 (2004).
[14] E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater, The need for a light and spectroscopy concept inventory for assessing innovations in introductory astronomy survey courses, Astron. Educ. Rev. **4**, 20 (2006).
[15] S. C. Bilici, F. O. Armagan, N. K. Cakir, and N. Yuruk, The development of an Astronomy Concept Inventory (ACI), Proc. Soc. Behav. Sci. **15**, 2454 (2011).
[16] K. Powell, Science education: Spare me the lecture, Nature (London) **425**, 234 (2003).
[17] J. Lazendic-Galloway, M. Fitzgerald, and D. H. McKinnon, Implementing a studio-based flipped classroom in a first year astronomy course, Int. J. Innov. Sci. Math. Educ. **24**, 35 (2016).
[18] P. M. Sadler, D. Haller, and E. Garfield, Observational journals: an aid to sky watching, J. Coll. Sci. Teach. **29**, 245 (2000).
[19] J. S. Farnsworth, L. Baldwin, and M. Bezanson, An invitation for engagement: assigning and assessing field notes to promote deeper levels of observation, J. Nat. Hist. Educ. Exp. **8**, 12 (2014).
[20] S. Lewis and C. Mills, Field notebooks: A student guide. J. Geo. Higher Educ. **19**, 111 (1995).
[21] J. G. Cromley, Y. Du, and A. P. Dane, Drawing-to-learn: does meta-analysis show differences between technology-based drawing and pen-and-pencil drawing?, J. Sci. Educ. Technol. **29**, 216 (2020).
[22] G. Deming and B. Hufnagel, Who's taking ASTRO 101?, Phys. Teach. **39**, 368 (2001).
[23] W. Bisard, R. Aron, M. Francek, and B. Nelson, Assessing selected physical science & earth science misconceptions of middle high through university preservice teachers, J. Coll. Sci. Teach. **24**, 3842 (1994).

[24] M. Zeilik, C. Schau, and N. Mattern, Misconceptions and their change in university-level astronomy courses, Phys. Teach. **36**, 104 (1998).
[25] M. Zeilik and W. Bisard, Conceptual change in introductory-level astronomy courses, J. Coll. Sci. Teach. **29**, 229 (2000).
[26] B. Hufnagel, T. Slater, G. Deming, J. Adams, R. L. Adrian, C. Brick, and M. Zeilik, Pre-course results from the Astronomy Diagnostic Test, Pub. Astron. Soc. Aust. **17**, 152 (2000).
[27] R. Trumper, University student's conceptions of basic astronomy concepts, Phys. Educ. **35**, 9 (2000).
[28] R. Trumper, Assessing students' basic astronomy conceptions from junior high school through university, Aust. Sci. Teachers J. **47**, 21 (2001).
[29] B. Hufnagel, Development of the Astronomy Diagnostic Test, Astron. Educ. Rev. **1**, 47 (2002).
[30] J. O'Byrne, Introductory astronomy: Setting goals and measuring achievements, in *Proceedings of The Australian Conference on Science and Mathematics Education*, edited by A. Fernandez (Camperdown, UniService Science, 2002), p. 57.
[31] E. V. Slater, J. E. Morris, and D. McKinnon, Astronomy alternative conceptions in pre-adolescent students in Western Australia, Int. J. Sci. Educ. **40**, 2158 (2018).
[32] http://www.physics.usyd.edu.au/super/ADT.html.
[33] M. Zeilik, C. Schau, and N. Mattern, Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters, Am. J. Phys. **67**, 923 (1999).
[34] http://www.flaguide.org/cat/diagnostic/diagnostic6.htm.
[35] L. Kohlhauf, U. Rutke, and B. Neuhaus, Influence of previous knowledge, language skills, and domain-specific interest on observation competency, J. Sci. Educ. Technol. **20**, 667 (2011).
[36] K. D. Forbus and S. Ainsworth, Editors' introduction: sketching and cognition, Topics Cogn. Sci. **9**, 864 (2017).
[37] C. Hoffman and B. Wittman, Introduction: Knowledge in the making: Drawing and writing as research techniques, Science in context **26**, 203 (2013).