Astronomy Education Journal, 2022, Vol. 02, No. 1

doi: 10.32374/AEJ.2022.2.1.043ra

ISSN: 2004-2981 • Published in Sweden

RESEARCH ARTICLE

A New Curriculum Development Model for Improving Undergraduate Students' Data Literacy and Self-Efficacy in Online Astronomy Classrooms

Molly N. Simon[1,*], Edward E. Prather[2], Isaac S. Rosenthal[3], Michael Cassidy[4], James K. L. Hammerman[4] and Laura Trouille[5]

[1]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA ; [2]Department of Astronomy, University of Arizona, Tucson, AZ, USA ; [3]School for the Environment, University of Massachusetts Boston, Boston, MA, USA ; [4]TERC, Cambridge, MA, USA ; [5]The Adler Planetarium, Zooniverse, Chicago, IL, USA, Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, Evanston, IL, USA

[*]molly.n.simon@asu.edu

Abstract

There is a critical need for research-based active learning instructional materials for the teaching and learning of STEM in online courses. Every year, hundreds of thousands of undergraduate non-science majors enroll in general education astronomy courses to fulfill their institution's liberal arts requirements. When designing instructional materials for this population of learners, a central focus must be to help learners become more scientifically and data literate. As such, we developed a new, three-part, curricular model that was used to inform the creation of active-learning instructional materials designed for use in online courses. The instructional materials were designed to help introductory astronomy students engage meaningfully with science while simultaneously improving their data literacy self-efficacy (especially as it pertained to making evidence-based conclusions when presented with a variety of data representations).

We conducted a pilot study of these instructional materials at nine different colleges and universities to better understand whether students' engagement with these materials lead to improved beliefs and self-efficacy. The results of our student survey analysis showed statistically significant changes on survey items that assessed students' beliefs about science engagement, citizen science, and their data literacy skills. Additionally, we assessed whether faculty who implemented these materials were able to easily incorporate them into existing online astronomy courses. The instructor feedback emphasized that our curriculum development model did successfully inform the creation of easy-to-implement instructional materials, generating the potential for widespread dissemination and use at the undergraduate level.

Keywords: Self-efficacy; Citizen science; Data literacy; Curriculum development; Higher education

1 Introduction

Every year, hundreds of thousands of non-science majors enroll in general education astronomy courses to fulfill their institution's liberal arts requirement (Rudolph et al., 2010). Upon graduation, these students go on to become our nation's teachers, business leaders, journalists, social media influencers, lawyers, historians, artists, politicians, as well as taxpayers, voters, and parents. Considering that these courses are often students' last formal exposure to science, it is critical to help them further develop ideas and skills that will allow them to grapple with contemporary issues and make meaningful contributions to humanity beyond the classroom. When designing instructional materials for this population of learners, a central focus must be to help learners become more scientifically and data literate (National Research Council, 1996). Increased data literacy skills serve non-STEM majors by empowering them to feel confident using quantitative reasoning in their everyday lives: e.g. when reading the news or voting on policy decisions that impact them and their communities (Feinstein et al., 2013). Furthermore, previous research has highlighted undergraduate students struggle with distinguishing between data and evidence (Lyons, 2011), and with making predictions, observations, or explanations when presented with real data (e.g., Kastens et al., 2009; Mattox et al., 2006; Tien et al., 2007). To this end, our goal was to create instructional materials that help introductory astronomy students improve their scientific self-efficacy (e.g., Bandura, 1977) as it pertains to analyzing data and making evidence-based conclusions when presented with a variety of data representations. Furthermore, we aimed to improve students' beliefs about engagement in science and about citizen science more generally, all while increasing their knowledge of relevant astronomical topics. The creation of such instructional materials by the authors was ultimately informed by a new curriculum development model that includes three distinct parts:

1. A tutorial-based introductory activity that affords students the opportunity to develop representational competence and essential background knowledge of the discipline.
2. A science investigation that empowers learners to explore real data from the forefront of active research in STEM and allows them to make contributions to the scientific community.
3. A data analysis activity that encourages students to engage in critical reasoning, while making evidence-based conclusions in pursuit of answers to contemporary science questions.

While we have had success with developing pencil and paper active-learning instructional materials that help introductory astronomy students develop their conceptual understandings and reasoning abilities associated with several key astronomy topics (Prather et al., 2004; Hudgins et al., 2006; Wallace et al., 2012, 2016), these materials were not explicitly designed with the goals of increasing students' data literacy self-efficacy or contributing meaningfully to science. To better accomplish these goals, we moved away from static pencil and paper activities to an online investigation that made it possible to provide students with access to authentic data and engaging analysis tools. The work described in this paper contributes to the growing need for evidence-based, active learning, instructional materials suited for the online classroom, and provides a framework for others who are engaged in designing online educational experiences intended to increase accessibility and engagement with data and where broadening participation in science is a priority.

Student enrollment in online courses has increased significantly in recent years due to the array of benefits they provide to students, instructors, and institutions alike (Cooper et al., 2019; Allen and Seaman, 2013). These courses are more accessible than in-person courses in that they can be completed by individuals from all over the world without having to commute to a physical campus. This attracts students who may otherwise face challenges attending courses in-person, including military personnel, international students, and working parents. Online courses are demonstrably successful at broadening participation in higher education, offering a pathway to make STEM education more inclusive and equitable as a result (e.g., Perera et al., 2017; Mead et al., 2020). Consequently, online courses have been adopted by nearly two-thirds of higher education institutions (Allen and Seaman, 2013). Although enrollment in these courses continues to increase, the development and widespread availability of learner-centered, research-based instructional materials explicitly designed to match the knowledge, beliefs, and abilities of our target population remains scarce. This scarcity has been highlighted even more profoundly during the COVID-19 pandemic which has forced courses generally taught in-person to be quickly transformed into online formats.

The remainder of this paper is structured as follows: first, we provide details into the motivation and theoretical framework that informed the design of our curriculum development model. We then unpack the different parts of our curriculum development model, highlighting how we integrated citizen science, unique discipline representations, and research-based active learning strategies into online instructional materials that effectively bring the analysis of big data into general education online classrooms. Next, we describe the methodology used, and results from, an initial pilot study of these instructional materials implemented in online general education astronomy courses. Finally, we provide an interpretation of our results and describe implications for future work.

2 Theoretical Framework Overview

Creating a set of active learning instructional materials to increase students' data literacy and self-efficacy for use online requires a curriculum development model that is informed by a theoretical perspective on how to effectively and efficiently intellectually engage learners in complex disciplinary ideas and representations. Our new curriculum development model has been strongly informed by theoretical work in social semiotics (how different groups of people create and maintain their own interpretations or meanings of relevant concepts and representations), how social semiotics influences learners' development of representational competence, and the role of transduction which involves "the ability to move between different types of semiotic systems, e.g. between a table and a graph (Volkwyn et al., 2019, 2020)."

Learners develop their representational competence at the beginning of our instructional materials so that they can engage in the level of data analysis and evidence-based reasoning the later parts of the materials are designed to foster. As such, we began the curriculum development process by examining a range of disciplinary representations traditionally used for our topics, documenting the different disciplinary meanings these representations afford, and how these meanings may be learned by our target population while in an online environment. Often, we find that traditional representations lack the disciplinary and pedagogical affordances that our curriculum will need to provide if our target population is to achieve representational competence. This necessitates that we develop new representations that present information in ways not typically found in the discipline. These representations are invaluable pedagogically because they feature highly stylized physical scenarios that depict distinct and unique discipline relationships (French and Prather, 2020). For this reason, these new representations are referred to as pedagogical discipline representations, or PDRs (Wallace et al., 2016). PDRs are depictions of discipline information with "specific, narrowly focused and well-understood disciplinary affordances". PDRs help students unpack and make connections between the ideas the PDR conveys, while also "enabling critical and disciplinary discernment (coming to recognize and understand what to focus on and interpreting it or making meaning using the appropriate context)" that ultimately allows learners to develop more robust and coherent mental models (French and Prather, 2020, p.2, and the references therein). Critical to our curriculum development model, which has a goal of increasing students' data literacy, was the creation of a new set of PDRs that requires students to analyze real data from the science community while simultaneously evaluating the physical properties of the phenomenon being studied. These new PDRs are designed to motivate online students to develop their understanding of the discovery process professional scientists go through when analyzing data to answer fundamental research questions.

Prior research efforts investigated how novice learners struggle with recognizing important information and relationships afforded by disciplinary representations and how learners engage in meaning-making for particular science contexts (e.g., Prather et al., 2004; Hudgins et al., 2006; Wallace et al., 2016). The findings from these efforts echo the assertions of French & Prather (2020) regarding novice learners, "they cannot yet critically discern the disciplinary affordances of multiple representations and coordinate them to make sense of disciplinary knowledge" (p.8). Each representation provides students with partial disciplinary understanding, but no individual representation can capture every aspect that the topic is intended to communicate (Fredlund et al., 2014). Multiple representations, however, may work in harmony to create a collective disciplinary affordance, offering a more complete understanding of the topic being investigated (Linder, 2013; French and Prather, 2020). Providing students with a variety of representations increases the probability that one or more of these representations will help students increase their representational competence. Further, getting students to engage in unpacking and discerning the meanings of multiple representations can significantly increase their disciplinary knowledge.

Our curriculum development model is designed to get students to engage in "disciplinary discourse," helping them engage with a variety of representations, selecting, interpreting, explaining, reflecting on and reconciling them, ultimately leading to increased disciplinary knowledge and reasoning ability. This requires that we couple multiple representations with a variety of intellectual tasks in ways that facilitate learning when moving from one semiotic system to another (transduction). To do this we employ a "variation approach to learning" (Linder and Fraser, 2006) perspective, in which we create task sequences coupled to data representations, (e.g., tables, drawings, and graphs) that provide our online learners access to a coherent set of disciplinary ideas, leading to increases in their data literacy and self-efficacy.

3 Unpacking the New Curriculum

In this section, we illustrate how the theoretical perspectives highlighted in the previous section have been used to inform the development of a set of instructional materials explicitly for use in general education college astronomy courses. We were motivated to provide instructional materials that could be delivered online, that are well-matched with topics instructors already teach, and that are modular—allowing them to easily fit into an existing online course without major time commitment or course modification. First, it is important to describe the science topic and citizen science project that are at the center of this investigation. Then, we unpack the flow of activities used to bring our theoretical perspectives and learning outcomes to fruition.

3.1 Citizen Science as a Vehicle for Analyzing Large Datasets

The instructional materials described in the remainder of the paper are centered around a citizen science project delivered on the Zooniverse platform (`https://www.zooniverse.org/`; (Lintott et al., 2008)) called Planet Hunters (`https://www.planethunters.org`), which, in its current iteration, has volunteers analyze data from the Transiting Exoplanet Survey Satellite (TESS) to identify potential new exoplanet candidates. Since the first discovery of an exoplanet (planet outside of our Solar System) in 1995 (Mayor and Queloz, 1995), the study of exoplanets has grown significantly, with over 5,000 confirmed exoplanets discovered as of June 8th, 2022 (NASA Exoplanet Archive, 2022). Studies of exoplanets and their properties have provided astronomers with powerful insights into the formation and evolution of planetary systems. Astronomers use a variety of detection methods when searching for extrasolar planets, but the transit method remains the most robust, accounting for nearly 75% of the exoplanets listed on the NASA Exoplanet Archive. Planetary transits occur when a planet passes in front of its host star, leading to a small, periodic, decrease in observed brightness of the host star. To identify potential exoplanet candidates, astronomers analyze transit light curves, which show a star's changes in brightness over time. To date the vast majority of planets discovered via the transit method have sizes and exist in locations that are not consistent with the planets in our Solar System. Astronomers continue to inquire (and our activity is centrally focused on) whether our Solar System's architecture is unusual amongst the billions of potential solar systems in our galaxy, or whether differences are due to the biases that exist in our current detection methods that favor the detection of large planets orbiting close to their host star(s).

The complete verification process for an exoplanet discovery is quite complex. The automated TESS planet detection data pipeline requires a minimum of at least two transits and a high signal–to–noise ratio to mark a potential detection. Earlier iterations of the Planet Hunters project demonstrated that human vetting of light curves can outperform the automated pipeline for specific types of transits: single, longer-period transits in particular (Eisner et al., 2021). This allows the citizen science community to make a truly meaningful impact in the transit (and potential planet) detection process. We incorporated the Planet Hunters investigation of exoplanet data into our instructional materials because it focuses on a compelling subject that is popular with instructors teaching our target courses. Planet Hunters also presents a low (technological) barrier to entry and requires no previous research experience, while promoting online students' active engagement with real data and offering them an opportunity to contribute to exoplanet research.

Studies of public participation in citizen science projects have measured increases in scientific and data literacy (e.g., Cronje et al., 2011; Crall et al., 2013), increases in confidence for learning and contributing to science (Masters et al., 2016; Greenhill et al., 2016), a stronger sense of place and connection to the environment (Toomey et al., 2020), increases in long-term interest in continued participation in research (Dickinson and Bonney, 2012), and positive shifts in participants' attitudes towards science and scientific beliefs (Price and Lee, 2013). However, the participants in many of these studies were middle-aged men who were intrinsically motivated to contribute to scientific research (Raddick et al., 2013), and as a result, these outcomes may not be generalizable to college students taking an online introductory science course in a formal education setting. These students are required to enroll in science courses to fulfill their institutions' liberal arts requirements and represent a much more diverse group of individuals whose demographic makeup is more representative of the general population. By introducing citizen science into our curriculum development model, we sought to determine whether these positive outcomes are generalizable to our more diverse population of students in a formal education setting.

3.2 The Curriculum Development Model in Context

There are three main parts (described in Section 1) to our curriculum development model and corresponding instructional materials, referred to hereafter as the 'Planet Hunters Activity.' The Planet Hunters Activity was developed to engage students in understanding the process and interpretation of data related to discovering exoplanets with the transit method. Below we unpack key representations and tasks from each of these three parts of our activity to demonstrate how our theoretical perspectives and citizen science approach work together to achieve our learning outcomes.

Part I of the Planet Hunters Activity is a Lecture Tutorial (Prather et al., 2004) that serves to situate the learner into the disciplinary context of the investigation. This part of the activity leads students to study a sequence of PDRs using a variety of critical reasoning tasks designed to help students develop their disciplinary knowledge associated with:

1. Orientation of the planet-star system with line of sight to Earth in order to observe transits
2. Planet size and light curve dip depth
3. Planet distance and orbital period
4. Determining whether there are multiple planets in a system
5. Comparing the properties of the planets in our Solar System with those of commonly discovered exoplanets

In Figure 1, we provide a PDR that anchors a control of variables activity found near the beginning of Part I, in which students are asked to determine which orbital properties shown in the drawings correspond with which properties shown in the light curves. This requires students to engage in transduction of information encoded in two different discipline representations. This allows the learner to gain valuable experience with unpacking and discerning the important information afforded by these central representations of the discipline. The remainder of Part I provides students with experience analyzing a light curve that models real data for a multi-planet system and requires students to make comparisons between exoplanetary systems and our own Solar System.

With actual exoplanet data, it is difficult to find a multi-planet system where each of the planets have clear and distinct transits, the planets are of different sizes, and where each exoplanet has more than one transit. However, our own prior teaching experience has uncovered that students need to analyze a light curve with these properties if they are to develop a model that they can use to understand more complex multi-planet systems. The light curve at the top of Figure 2 shows a unique PDR that was created by combining actual data from two separate exoplanetary systems to create a light curve that models a two-planet system containing exoplanets with different sizes and orbital periods, with easy to distinguish dips that repeat in the duration of the observation time shown in the graph. At the bottom of Figure 2, are two planetary system representations, one of which students have to identify as matching a presented light curve. Additionally, we ask students to make predictions about future transit events for this system by extrapolating the data from this light curve. This is what makes PDRs so powerful, they can create a scenario for students to consider that has all the right attributes to facilitate learning, which the actual exoplanet data cannot (easily) replicate.

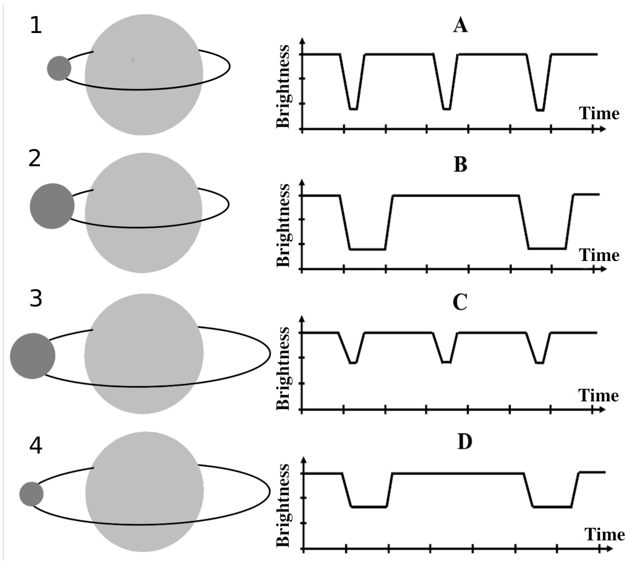

Figure 1. In this pedagogical discipline representation (PDR), students are required to match the planetary system to its corresponding transit light curve.

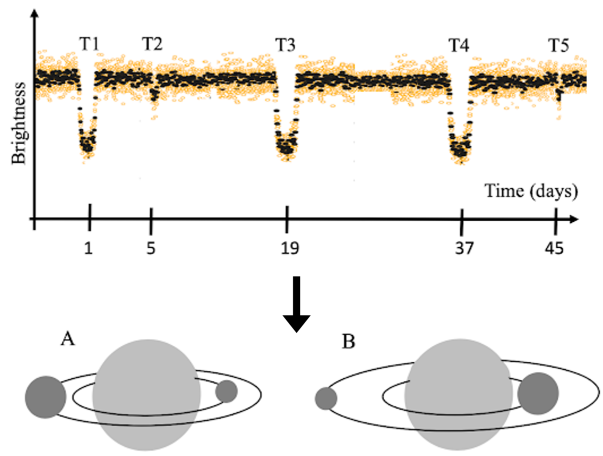

Figure 2. A multi-planet system PDR that was created through the amalgamation of two independent light curves. Students once again are required to match the light curve to a representative planetary system, and derive planetary characteristics from the light curve.

The current dataset for transiting exoplanets contains many large planets at locations closer than Mercury (NASA Exoplanet Archive, 2022). Astronomers did not initially expect to find large planets with such close-in orbits, since early planet formation models were developed based on our knowledge of our Solar System. Helping students to develop explanatory models to account for these planets, and the notion that our Solar System is perhaps atypical, are goals of this investigation. Therefore, we felt it important, at this point in the activity, to scaffold their understanding by introducing planets with these specific characteristics in Figure 2. Students are also provided with a data table containing the radii and orbital periods of the planets in

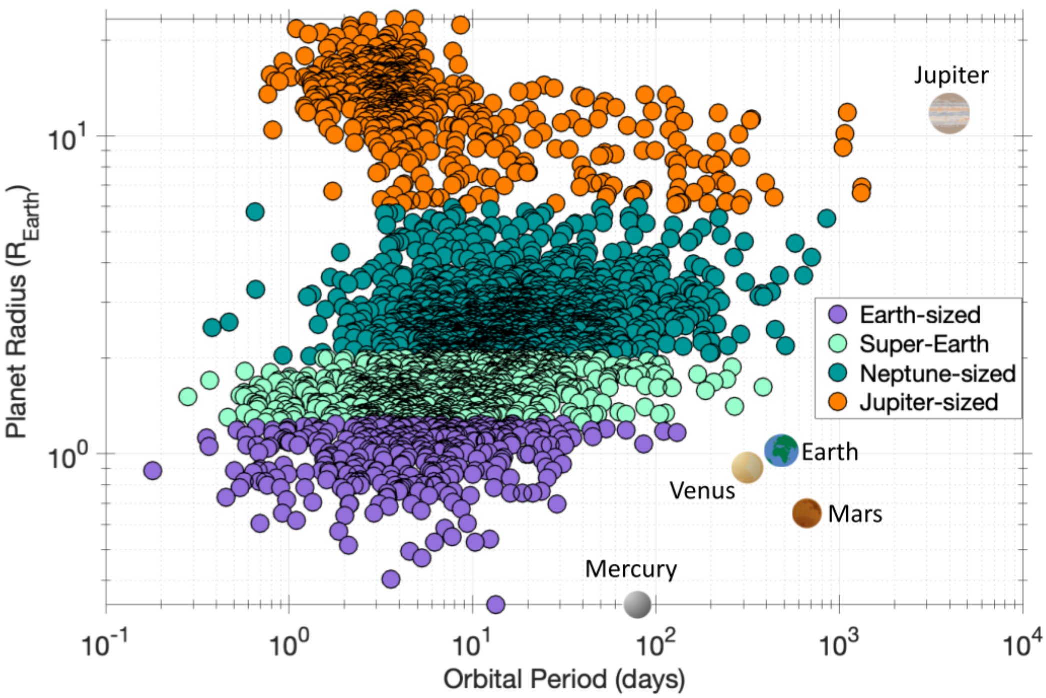

Figure 3. Planet radius versus orbital period for four types of exoplanets (data from the NASA Exoplanet Archive). The Solar System's planets are plotted for reference.

our Solar System. Students are asked to evaluate whether the data presented in the light curve from Figure 2, could represent transits for any pair of planets in our Solar System. This PDR was explicitly created such that the orbital periods depicted place both planets closer than Mercury, with one planet being quite large (modeling a Jovian planet) and therefore neither planet would be consistent with the planets in our Solar System.

In Part II of the activity, students are asked to apply their understanding of exoplanet detection and analyze real data from the TESS mission via Zooniverse's Planet Hunters project, and in doing so directly contribute to an active research investigation. Before analyzing the most recent TESS light curve data, students complete a short training module in which they work through a curated data set chosen to help them develop proficiency with the tools used to identify dips in light curves. In creating our training module, we employed aspects of game design (Boller and Kapp, 2017) to maintain students' attentiveness and motivation and increase the likelihood that they would be able to correctly identify transits when analyzing real data from TESS on their own. The module begins with a gif highlighting what successful and unsuccessful moves for the identification of dips would look like, and which features of the light curve they should focus on. Additionally, students receive immediate feedback on whether or not they had properly identified the transits in each individual light curve before moving on to the next one. After completing this training module, students are instructed to analyze 15–20 light curves, contributing to the research efforts of the Planet Hunters project (from the first implementation of our new activity, there were over 15,000 classifications completed by students).

At the end of Part II, students reflect on whether their experience in the training module helped them as they encountered the authentic, albeit more complex data, from the TESS mission. Additionally, they are asked to reflect on whether the detection of transits is common. Students were typically quick to note that the real TESS dataset does not contain abundant transits, demonstrating that transit events are rare. Further, students found the data analysis activities using the PDRs from Part I helpful when analyzing the more subtle and complex data they encountered in Part II. Having students stop and reflect on these experiences may lead to improvements in their beliefs about contributing to and doing science, something many of these students previously thought was inaccessible to them (e.g., Cera et al., 2013).

In Parts I and II of this activity, our main objectives were to help students develop their knowledge and skills regarding analyzing data from transit events and how to interpret the physical characteristics of the exoplanetary systems that these light curves represent. The overarching goal of Part III, however, is to provide students with a robust data analysis experience that helps them to determine whether our Solar System is typical or unique amongst the planetary systems we find near us in the galaxy. This process mirrors the scientific community's evidence-based reasoning process about a physical system and uses the same datasets used by professional astronomers, further supporting students to see themselves as capable scientific contributors. To this end, in Part III students are introduced to a series of data-focused PDRs (generated from the NASA Exoplanet Archive) containing all exoplanets confirmed via the transit method as of June 2020. Students examine exoplanet radius and orbital period histograms (and summary tables) independently and compare these properties directly to a table with corresponding properties for our Solar System's planets. They discover that there are many exoplanets the size of Earth, that there is a class of exoplanet that does not exist in our Solar System (Super-Earths, planets with masses greater than Earth but less than Neptune), that the majority of all detections have been of planets larger than Earth with orbital periods that would place them within the orbit of Mercury, and that it is easiest to detect exoplanets that are large, and near their host star.

In prior years teaching this topic, we observed that while students can reason about exoplanet radius or exoplanet orbital period independently, they struggle to draw conclusions or make predictions about exoplanetary systems involving the relationship between these two variables. As such, we designed a PDR that is an amalgamation of all the exoplanet data students had previously encountered along with our Solar System's planets plotted on the same figure (Fig. 3). This PDR has a high pedagogical affordance as it provides our learners the opportunity to unpack and discern the interplay among the relevant variables needed to understand transiting exoplanets.

An important aspect of our curriculum development model is to purposefully include places in the learning sequence where students engage in metacognition, and evaluate how data is used to answer the question at the center of this investigation. This is exemplified in the question sequence following Figure 3. First, students are asked whether they would update or change their prior responses given the data shown in this new representation. Next, students evaluate whether the distances from the Sun and sizes of the planets in our Solar System serve as a good representation of the overall characteristics for the exoplanets discovered using the transit method.

The data, representations, and task sequence used in this activity are likely to lead students to the conclusion that the planets of our Solar System are a poor or incomplete representation for the exoplanets we are discovering with the transit method. Although this perspective may be supported by current transit data, there are several constraints unique to the transit method that led to biases in the data and limit the inferences students can make about the abundances of different types of exoplanets. To bring awareness to these constraints and biases and have students evaluate these ideas, we use a pedagogical strategy that asks them to analyze a hypothetical student discussion.

One approach that we use to help students address lingering scientifically inaccurate ideas they may still have regarding a particularly difficult topic is to model a hypothetical student debate (McDermott et al., 1998). This technique presents students with a conversation between 2-3 hypothetical classmates written in plain language that sounds similar to how undergraduates would realistically speak to each other. Students are then tasked with selecting and defending their position in the debate by providing an explanation regarding which hypothetical classmate they agree with and why. From a learning perspective, these mock debates provide students with the opportunity to reflect on, challenge, and ultimately address any potential conceptual and reasoning difficulties, which helps mediate meaningful and lasting conceptual change (Posner et al., 1982; Prather et al., 2004).

> *Student 1: Based on the histograms and the graph in Figure 3, it's clear that the transit method is better at finding larger planets that are close to their stars, so it's no wonder we have a data set that looks very different from the locations and sizes of planets in our Solar System.*
>
> *Student 2: Maybe all we need to do is search for a longer time, and we will start to find more large planets far away on long period orbits. This will show us that Jupiter-sized planets far away are more common than the current data set suggests.*
>
> *Student 3: Whether our Solar System is typical or not, this data has shown us that there is a new category of exoplanet that we don't see in our Solar System, and that all types of planets can be found closer to their stars than what we previously thought.*
>
> *Do you agree or disagree with any/all of these students? Explain your reasoning.*

In this discussion, students articulate three different perspectives on the data, methodology and interpretations of transiting exoplanets and the comparisons that can be drawn with the planets of our own Solar System. The language and arguments presented in this hypothetical discussion models the same non-technical language our students use, as opposed to how a set of experts might defend their ideas. Through this discussion we afford students the opportunity to address the limitations of their conclusions, and to hypothesize how advances and future developments in exoplanet detection methods may lend to the discovery of planets with characteristics more in line with those in our Solar System's planets.

The final sequence of representations and tasks in Part III allow students to make comparisons between data for exoplanet transits that were first identified by citizen scientists with the larger data set of transiting exoplanets that have been confirmed by experts in the scientific community. Students find several similar exoplanet characteristics from analyzing the data from citizen scientists and from the larger scientific community. This result is leveraged in the final question which asks whether they would support the assertion that citizen scientists are making valuable contributions to the discovery of exoplanets.

4 Methods

We conducted a pilot study with students in courses at nine institutions of higher education during the 2020–2021 academic year. Additionally, we conducted interviews with pilot course instructors to gain insight into how they integrated the Planet Hunters Activity into their already established, undergraduate, newly online astronomy courses. We report here the preliminary results of these assessment efforts.

4.1 Setting and Participants

The Planet Hunters Activity was tested in 10 general education astronomy courses at nine institutions of higher education during the Fall and Spring of the 2020–2021 academic year. The testing institutions included community colleges and four-year colleges and universities with varying degrees of research emphasis. The Planet Hunters Activity was implemented as a 75–90 minute intervention in each of these courses. Due to the COVID-19 pandemic, eight of the participating courses implemented the activity synchronously via Zoom, and two courses chose to offer the activity fully asynchronously. A complete list of participating institutions is provided in Table 1.

The students enrolled in the 10 aforementioned courses were predominantly undergraduate non-science majors taking an introductory astronomy course to fulfill their institution's general education requirements. Students were commonly in the first 2 years of their undergraduate tenure, and the demographic makeup of these courses are typically consistent with the institution's undergraduate population more broadly. General education science courses can enroll large numbers of students, especially at public universities, so enrollments in these courses can range from 15 to upwards of 200 students depending on the institution type. To conduct research with students and course instructors alike, the research team required approval from the institutional review board (IRB) of all participating institutions where data are being analyzed. This multi-site study has been approved by all required institutions and has been classified as "exempt," meaning the project does not pose any harm to the study participants and is not subject to further review unless there are significant changes made to the study protocol [1].

4.2 Assessments

Our evaluation of the effectiveness of the Planet Hunters Activity focused on assessing students' data literacy self-efficacy and beliefs about science engagement and citizen science. These were the main goals that our curriculum development

1 Engaging Non-Majors in Classroom-Based Citizen Science Experiences, Arizona State University (IRB of Record) ID: STUDY0001215

Table 1. A list of institutions and the total number of students that participated in the pilot testing efforts during the 2020–2021 academic year. An additional 10 students neglected to report an institution on their survey responses.

Institution	Institution Type	Number of Student Participants
Arizona State University	Public University, Very High Research Activity (R1)	229
University of Colorado, Boulder	Public University, Very High Research Activity (R1)	199
University of Minnesota, Twin Cities	Public University, Very High Research Activity (R1)	479
American University	Private University, High Research Activity (R2)	20
University of Alaska, Anchorage	Public University	15
University of North Carolina, Asheville	Public Liberal Arts University	48
College of Idaho	Private Liberal Arts College	23
Guilford Technical Community College	Public Community College	67
Mt. San Antonio College	Public Community College	19
		Total = 1099

Table 2. A complete list of items from the student survey. Items 4-10 appeared on both the pre- and post-tests, and are grouped into their respective factors.

Items	Category/Factor
1. I usually understand concepts taught to me in my science classes.	Pre-test only
2. Science plays an important role in our society.	Pre-test only
3. I trust the results that come from scientific research.	Pre-test only
4. I want to make contributions to science that I find meaningful.	Science Engagement
5. I can make contributions to science that I find meaningful.	Science Engagement
6. I understand how data and evidence can be used to inform scientific conclusions.	Data Literacy Self-Efficacy
7. I am comfortable using data and evidence to inform my own scientific conclusions.	Data Literacy Self-Efficacy
8. I am confident in my ability to use data representations (graphs, tables, and charts) when seeking out answers to questions.	Data Literacy Self-Efficacy
9. Participation in citizen science would allow me to make meaningful contributions to the scientific community.	Citizen Science
10. Citizen science projects can make valuable contributions to scientific research.	Citizen Science
11. This activity made me more likely to classify on Planet Hunters Again.	Post-test only
12. This activity made me more likely to go to the Zooniverse website and explore other citizen science projects I can contribute to on my own.	Post-test only
13. This activity improved my ability to understand how data and evidence are used to inform scientific conclusions.	Post-test only
14. I would look forward to doing another citizen science-based activity again in my class.	Post-test only

model aimed to support. As such, we developed a 14-item, Likert-style survey loosely inspired by Estrada-Hollenbeck et al. (2011) and Coburn (2001). Each of these items were rated on a scale of 1 (strongly disagree) to 7 (strongly agree). Three of the fourteen items were pre-test only, and four items were only asked as part of the post-test, as they focused more specifically on students' perceptions of the activity's potential impacts. The remaining seven items were asked in a pre/post-test fashion, at the beginning of the semester (pre) and then again within a week after the activity was implemented in the classroom (post).

The seven items that appeared on both the pre- and post-tests were grouped into the following three factors:

1. Science Engagement (students' beliefs about science engagement)
2. Data Literacy Self-Efficacy (ability to make evidence-based conclusions when presented with data/data representations)
3. Citizen Science (students' beliefs about citizen science)

For a complete list of survey items (and their subsequent factor groupings), see Table 2.

To determine the validity of our proposed groupings, we performed both an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). We used the following indices and thresholds to assess model fit: RMSEA $\leq$ 0.06, 90% CI $\leq$ 0.06, SRMR $\leq$ 0.08, CFI $\geq$ 0.95, and TLI $\geq$ 0.95 (Hu & Bentler, 1999; Brown, 2015). A three-factor model was selected based on visual inspection of the scree plot and consideration of the goals of our three-part curricular model. CFA testing of this three-factor model showed acceptable fit (RMSEA = 0.059, 90% CI = 0.028, SRMR = 0.034, CFI = 0.98, and TLI = 0.96). For more details regarding these analyses, refer to Appendix A.

Internal reliability of our survey was assessed using Cronbach's alpha, with a threshold of 0.7 indicating adequate reliability (Cronbach, 1951). Each latent variable (factor) in the pre/post-dataset was above the threshold for internal consistency (Citizen Science = 0.74, Data Literacy = 0.73, Science Engagement = 0.76). Furthermore, the pre-test only items had an α = 0.76, and the post-only items had an α = 0.82.

In addition to the student survey, we aimed to understand how instructors implemented our activity in their classroom. Each pilot course instructor and their teaching assistants (N = 17) participated in an end–of–semester exit interview with the external evaluation team at TERC, a non-profit STEM education research and development company in Cambridge, MA. Course instructors were asked a series of questions regarding the activity's ease–of–use, as well as the perceived value of implementing a data-driven, citizen-science based activity in their courses. A summary of the instructor feedback can be found in Section 5.2.

5 Results and Discussion

Table 3. Student survey participant distribution after cleaning the data as described in Section 5.1

Administration	Number of Students (N)	Percentage of Total (N/N_total)
Pre-test	954	**86.8**
Post-test	480	**43.7**
Matched pairs	325	**29.6**

5.1 Student Data

The data reported here is from the student survey administered before and after completion of the Planet Hunters Activity. Each student received the same version of the survey, which was administered online via Qualtrics `https://www.qualtrics.com/`. Before the data were analyzed, we removed any students from the sample who finished the survey in less than 30 seconds, as this indicated that they did not take the time to provide earnest responses to our survey. Next, we removed students from the dataset who left more than two items unanswered to avoid including incomplete responses. Finally, students' pre/post responses were matched (matched pairs) when possible. The final number of student responses after the data cleaning process can be found in Table 3. The high level of attrition between pre- and post-survey responses, and relatively small fraction of matched pairs data shown in Table 3 can be attributed to a variety of factors such as: the added stress placed on students during the COVID-19 pandemic, students dropping the course, students not participating in or attending class on the day of either pre- or post-testing, or students opting out of the surveys due to their negligible impact on students' course grade (the surveys were marked for completion and ungraded).

In the case of our student survey data, the matched pairs data is the most robust, as it is the only subset of our data where we can ensure that the students took the pre-test, completed the activity, and then took the post-test. To determine whether the matched pairs dataset was representative of the entire population of students surveyed, we performed an unpaired Wilcoxon Ranked Sum test (due to the ordinal, non-normal nature of our data) between the un-matched pre-test responses and the matched pre-test responses. Normality was assessed with both Mardia's multivariate kurtosis test and Mardia's multivariate skewness test, indicating significantly non-normal data ($p < 0.01$ in all cases). We found no statistically significant difference between the two populations ($Mean_{unmatched} = 5.66$, $Mean_{matched} = 5.67$, Z = -0.02, $p = 0.984$, df = 953). We performed the same statistical test between the un-matched and matched post-test data and once again found no statistically significant difference between the two populations ($Mean_{unmatched} = 5.62$, $Mean_{matched} = 5.79$, $Z = -1.78$, $p = 0.075$, df = 478). As such, we use the matched pairs data for our analysis of student survey results given that this enables the use of pairwise statistical analyses.

Within the student survey data, we were most interested in quantifying the change between students' survey responses for the items that appeared on both the pre- and post-tests (Items 4-10 in Table 2). As a first step, we computed the mean and standard deviation of items 4-10 on both the pre- and post-tests for each student in the matched dataset. We then performed a Wilcoxon Signed Rank Test to compare the pre- and post-means. For the matched pairs data, the results from the pre-test ($Mean_{pre} = 5.56$) and post-test ($Mean_{post} = 5.69$) comparison indicated that engagement with the Planet Hunters Activity had a significant positive impact on student students' beliefs and self-efficacy ($Z = -5.262$, $p = <0.01$) overall.

As an additional layer of analysis, we explored the differences between students' pre- and post-test responses for each of the three factors described in Section 4.2. To do this, we calculated the mean and standard deviation of student responses to the survey items within each factor. Again, we used a Wilcoxon Signed Rank Test to compare pre- and post- means for each of the factors. The results can be found in Table 4.

Overall, we found that students responded positively to the items on both the pre- and post-surveys (including the three pre-test only items). Considering that the pre-test item averages were high to begin with, we did not expect to see dramatic differences between the pre- and post-tests. This is not uncommon for self-efficacy-type surveys, where students notoriously respond with high positive values to Likert scale-style questions (Wallace et al., 2013). Although the effect sizes were understandably small, we did see statistically significant ($p < 0.05$) positive increases for all three of our factors. These results are particularly encouraging considering our curricular model (used to inform the development of the Planet Hunters Activity) placed specific emphasis on improving students' beliefs and self-efficacy across three specific domains (Table 4). As highlighted in Section 3.2, Part II of Planet Hunters Activity focused on providing students with a citizen science investigation that would allow them to meaningfully participate in active science research using real data. Additionally, Parts I and III of the Planet Hunters Activity focused on engaging this population with novel data representations and directed tasks that were explicitly created to elevate students' confidence surrounding their data literacy skills. Although preliminary, our student survey data positively supports our initial curriculum development goals.

In addition to the items that appeared on both the pre- and post-tests, we included a set of four items that were only given to students on the post-test. The breakdown of student responses can be found in Figure 4 and includes only the matched-pairs dataset (N = 325). We used Wilcoxon Rank Sum tests to compare matched and unmatched student responses for each of these four items to ensure the matched dataset was representative of the entire sample of post-test responses. In every instance, we found no statistically significant difference between the matched dataset and the unmatched dataset for the post-test only items ($P > 0.05$ in all cases).

For the matched dataset, we found that students were much more likely to provide a positive response (76.6% positive versus 8.3% negative) to the prompt inquiring whether they would look forward to doing another citizen science activity in their class. The vast majority of students (89%) responded positively to the prompt that the activity improved their understanding of how "data and evidence are used to inform scientific conclusions." Furthermore, roughly 70% of students responded that after completion of this activity, they were likely to classify on the Planet Hunters project again, and to visit the Zooniverse website to explore other citizen science projects beyond what was required for this activity. These post-only item results bolster our aforementioned findings regarding the statistically significant improvements we observed in each of our three factors. Overall, the post-test only items suggest that our curriculum development model can inform the creation of instructional materials that help increase students' data literacy self-efficacy, while encouraging them to meaningfully engage in and contribute to active science research.

5.2 Instructor Interview Data

As described in Section 4.2, interviews with the college faculty and teaching assistants (N = 17) who used the Planet Hunters Activity in their courses were conducted by the external evaluators from TERC. The evaluation team aimed to better understand

Table 4. Results from factor-level pairwise Wilcoxon Signed Rank tests using the matched pairs data (N = 325). Mean, standard deviation (SD), z-scores, statistical significance (p <0.05), and Wilcoxon effect sizes are reported for the 3 factors described in Section 4.2. The mean and standard deviations were on a scale from 1 (strongly disagree) to 7 (strongly agree).

Factor/Category	Item Numbers	Pre-test Mean ± SD	Post-test Mean ± SD	z-score	p-value	Effect Size
Science Engagement	4, 5	5.08 ± 1.36	5.30 ± 1.24	-4.12	<0.001	0.10 (small)
Data Literacy Self-Efficacy	6, 7, 8	5.84 ± 1.00	5.91 ± 0.02	-2.28	0.023	0.05 (small)
Citizen Science	9, 10	5.62 ± 0.95	5.78 ± 0.99	-3.64	<0.001	0.11 (small)

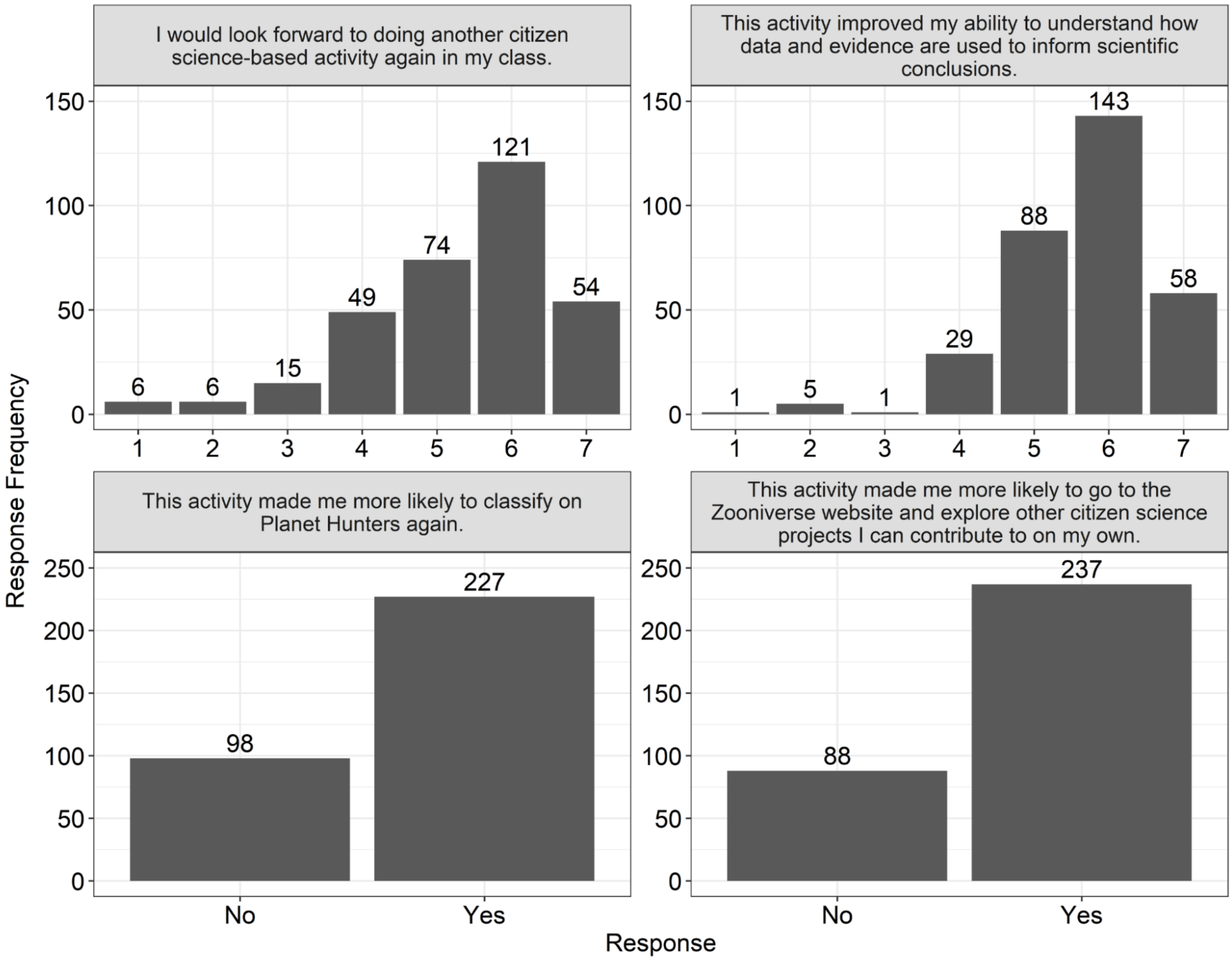

Figure 4. Bar chart of matched student responses (N = 325) to the four post-test only items. The top panel's x-axis can be interpreted as 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Neither Agree nor Disagree, 5 = Somewhat Agree, 6 = Agree, 7 = Strongly Agree.

instructors' experiences with using the Planet Hunters Activity, and what value there was in incorporating this activity into their courses. Interviews ranged from 16 to 60 minutes, averaging approximately 44 minutes, and were conducted through Zoom. Interviews were audio-recorded and transcribed. Transcripts were read and coded multiple times, looking for information related to major themes of the evaluation questions.

The external evaluation team found that instructors overall had positive experiences with the Planet Hunters activity across a wide range of respondents and higher institution types. Faculty felt the activity supported their goals for students; engaging with data and data visualizations, providing an experience with doing science, teaching some relevant content and, to some extent, motivating further participation with citizen science. The following quotes are from various instructors about how implementing the new instructional materials engaged their students and supported their course goals:

1. "Being able to see [and] analyze the data and help with the entire research analysis process. Students were very interested in that, and appreciated the ability [to]... that it was, you know, this is real data. This is a real research project."
2. "A lot of this really is like looking at what astronomers, like what does astronomy data look like? What do astronomers do, right? And so I think that a lot of people, they see these beautiful pictures on the cover of Astronomy textbooks or whatever, and they think that we take these beautiful photos of the sky, take them down to our desk, and just stare at them and write a paper. And so, trying to give them better visuals for what an astrophysicist is actually looking at, and you know, numerous different types of data that astronomers encounter. They were surprised."
3. "One of the reasons why I wanted to do this was to have students have a chance to actually experience and play with real data, and also participate in a citizen science project, because that's something that they can go on and do like themselves outside of class. So those two things, as well as

becoming more adept at looking at a graph and interpreting it. That's I think one of the big things, so data science literacy."

4. "One of the things is that since we had to move online, I was looking for things that would be easier to implement online, and was delighted that this would be an option. Exoplanets is a lab that is in our curriculum anyway. And so, it was easy to swap this in, versus the usual lab that we do, which is more of a pencil and paper worksheet, you know, like they don't get to interact with anything. It's just, look at some light curves. And so, I was looking for something more interactive, and this was really great."

The general consensus among instructors was that the activity was seen as high quality and easy to integrate into their courses, with the majority of the instructors reporting that they did not experience any challenges. Difficulties that were brought up by instructors included the canonical issues of online teaching, in particular, the struggles of getting students to pay attention and remain engaged when they are learning remotely.

During the interview, instructors were asked and encouraged to provide feedback on the instructional materials. Many of those that did not encounter challenges did not have any recommendations. However, some provided feedback on the content and how it was presented. These instructors' recommendations were to provide more data on the actual Planet Hunters project interface. Others commented on the phrasing and presentation of some questions and graphs which created confusion for students. Any changes to the instructional materials that came from our pilot testing were minor and are reflected in the version of the materials available for public use on the Zooniverse classroom webpage `https://classroom.zooniverse.org/#/`.

To conclude the interview, instructors were asked if they had any final thoughts. Below are two quotes that are representative of the positive sentiments we heard from instructors, and highlight how citizen science (and the Zooniverse, more specifically) can be used as a vehicle to bring data driven science to a large, diverse population of learners in an accessible way:

1. "Well, there's not enough time for me to say all the good things that I could say about Zooniverse. I think the benefit to the community, just the broader public, has been enormous. So I think these activities are fantastic, and sharing them, not only with colleges, but with high school and middle school educators, I think would be really beneficial. They're fantastic."
2. "Using it for the lab this semester, I had heard of things like it, but it was the first time that I had ever actually used something like Zooniverse for citizen science, and I've gotten sucked into it since. I've gotten actively involved with several of the other projects that are on Zooniverse doing the classifications, interacting in the discussion pages, and just doing that because I enjoy it and I think it's really interesting, and making science readily accessible for people that don't have extensive backgrounds in those fields."

Overall, instructors utilizing our activity were satisfied with the implementation in their courses, and it worked well in a variety of classroom contexts across many institution-types. Instructors were passionate about engaging non-science majors in citizen science because it provided opportunities for students to engage with real data and to contribute to an active research project. The instructor feedback further emphasized that our curriculum development model can successfully inform the creation of instructional materials that are easy to implement in existing courses, creating the potential for widespread dissemination and use. Perhaps most importantly, the pilot instructors recognized that our Planet Hunters Activity had the ability to broaden participation in science by making research accessible to a population of students that may have otherwise viewed contributing to science as something beyond their current skill set.

6 Conclusions and Future Work

In this paper, we have unpacked a curriculum development model that can be used to create instructional materials that meaningfully engage learners enrolled in college-level, general education astronomy courses taught online. The new Planet Hunters Activity derived from this model supported instructor course goals, was easy to implement in their pre-existing astronomy courses, and was well matched to the knowledge and abilities of the novice learners enrolled in these courses. Working through our instructional materials had a positive effect on students' attitudes and beliefs as they related to interpreting a variety of data representations and making meaningful contributions to science.

This work has offered a lens into the following future research directions:

1. We are currently conducting a study of a new set of instructional materials designed using the same curriculum development model to teach students about climate change, to better understand whether the results from this study are achieved in a different science content area. This study also includes a more extensive analysis of how our results may vary with regard to different demographic populations, institution-types, and teaching modalities.
2. We plan to design new assessments that allow us to directly measure student learning with regard to disciplinary content knowledge and students' ability to analyze data before and after completing the Planet Hunters Activity (following a procedure similar to what is outlined in Simon et al. 2019 and the references therein).
3. Lastly, we would like to better understand the nuances of successfully facilitating active learning in the online classrooms. While we have a detailed understanding of the different successful implementation techniques instructors use when facilitating active learning in in-person classrooms, we have significantly less insight into the analogous pedagogical choices and actions instructors incorporate in online courses.

Ultimately, this study bolsters the idea that online STEM courses can successfully attend to many of the important and diverse goals we have for our general education students. The curriculum development model described in this paper can ultimately serve as a guiding framework for the creation of future data-driven investigations beyond the discipline of astronomy.

7 Acknowledgements

The authors would like to thank Dr. Sanlyn Buxner (University of Arizona) and Dr. Chris Mead (Arizona State University) for helpful conversations regarding the statistical approach carried out in this work. M. N. Simon would like to thank graduate student Haylee Archer (Arizona State University) for her assistance updating several of the figures in this paper. Finally, the authors would like to thank each and every pilot instructor who implemented the Planet Hunters Activity in their classrooms. Without all of you, this study would not have been possible.

8 Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1821319.

References

Allen, I. E. and Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States.* ERIC.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191.

Boller, S. and Kapp, K. (2017). *Play to learn: Everything you need to know about designing effective learning games*. Association for talent development.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Cera, R., Mancini, M., and Antonietti, A. (2013). Relationships between metacognition, self-efficacy and self-regulation in learning. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*, 4(7):115–141. https://doi.org/10.7358/ecps-2013-007-cera.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational evaluation and policy analysis*, 23(2):145–170. https://doi.org/10.3102/01623737023002145.

Cooper, K. M., Gin, L. E., and Brownell, S. E. (2019). Diagnosing differences in what introductory biology students in a fully online and an in-person biology degree program know and do regarding medical school admission. *Advances in Physiology Education*, 43(2):221–232. https://doi.org/10.1152/advan.00028.2019.

Costello, A. B. and Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1):7. https://doi.org/10.7275/jyj1-4868.

Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J., and Waller, D. M. (2013). The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Public understanding of Science*, 22(6):745–764. https://doi.org/10.1177/0963662511434894.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334. https://doi.org/10.1007/BF02310555.

Cronje, R., Rohlinger, S., Crall, A., and Newman, G. (2011). Does participation in citizen science improve scientific literacy? a study to compare assessment methods. *Applied Environmental Education & Communication*, 10(3):135–145. https://doi.org/10.1080/1533015X.2011.603611.

Dickinson, J. L. and Bonney, R. E., editors (2012). *Citizen Science*. Cornell University Press, Ithaca, NY. https://doi.org/10.7591/9780801463952.

Eisner, N. L., Barragán, O., Lintott, C., Aigrain, S., Nicholson, B., Boyajian, T. S., Howell, S., Johnston, C., Lakeland, B., Miller, G., et al. (2021). Planet hunters tess ii: findings from the first two years of tess. *Monthly Notices of the Royal Astronomical Society*, 501(4):4669–4690. https://doi.org/10.1093/mnras/staa3739.

Estrada, M., Woodcock, A., Hernandez, P. R., and Schultz, P. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of educational psychology*, 103(1):206. https://doi.org/10.1037/a0020743.

Feinstein, N. W., Allen, S., and Jenkins, E. (2013). Outside the pipeline: Reimagining science education for nonscientists. *Science*, 340(6130):314–317. https://doi.org/10.1126/science.1230855.

Fredlund, T., Linder, C., Airey, J., and Linder, A. (2014). Unpacking physics representations: Towards an appreciation of disciplinary affordance. *Physical Review Special Topics-Physics Education Research*, 10(2):020129. https://doi.org/10.1103/PhysRevSTPER.10.020129.

French, R. S. and Prather, E. E. (2020). From a systematic investigation of faculty-produced think-pair-share questions to frameworks for characterizing and developing fluency-inspiring activities. *Physical Review Physics Education Research*, 16(2):020138. https://doi.org/10.1103/PhysRevPhysEducRes.16.020138.

Greenhill, A., Holmes, K., Woodcock, J., Lintott, C., Simmons, B. D., Graham, G., Cox, J., Oh, E. Y., and Masters, K. (2016). Playing with science: Exploring how game activity motivates users participation on an online citizen science platform. *Aslib Journal of Information Management*, 68(3):306–325. https://doi.org/10.1108/AJIM-11-2015-0182.

Hu, L.-t. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1):1–55. https://doi.org/10.1080/10705519909540118.

Hudgins, D. W., Prather, E. E., Grayson, D. J., and Smits, D. P. (2006). Effectiveness of collaborative ranking tasks on student understanding of key astronomy concepts. *Astronomy Education Review*, 5(1):1–22. http://dx.doi.org/10.3847/AER2006001.

Kastens, K. A., Agrawal, S., and Liben, L. S. (2009). How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-d geological structure. *International Journal of Science Education*, 31(3):365–393. https://doi.org/10.1080/09500690802595797.

Linder, C. (2013). Disciplinary discourse, representation, and appresentation in the teaching and learning of science. *European Journal of Science and Mathematics Education*, 1(2):43–49. https://doi.org/10.30935/scimath/9386.

Linder, C. and Fraser, D. (2006). Using a variation approach to enhance physics learning in a college classroom. *The Physics Teacher*, 44(9):589–592. https://doi.org/10.1119/1.2396777.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey*. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.x.

Lyons, D. J. (2011). *Impact of Backwards Faded Scaffolding Approach to Inquiry-Based Astronomy Laboratory Experiences on Undergraduate Non-Science Majors' Views of Scientific Inquiry.* Ph.D. Dissertation: University of Wyoming.

Masters, K., Oh, E. Y., Cox, J., Simmons, B., Lintott, C., Graham, G., Greenhill, A., and Holmes, K. (2016). Science learning via participation in online citizen science. *JCOM: Journal of Science Communication*, 15(3). https://doi.org/10.22323/2.15030207.

Mattox, A. C., Reisner, B. A., and Rickey, D. (2006). What happens when chemical compounds are added to water? an introduction to the model-observe-reflect-explain (more) thinking frame. *Journal of Chemical Education*, 83(4):622. https://doi.org/10.1021/ed083p622.

Mayor, M. and Queloz, D. (1995). A jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359. https://doi.org/10.1038/378355a0.

McDermott, L. C., Shaffer, P. S., et al. (1998). *Tutorials in introductory physics*. Prentice Hall Upper Saddle River, NJ.

Mead, C., Supriya, K., Zheng, Y., Anbar, A. D., Collins, J. P., LePore, P., and Brownell, S. E. (2020). Online biology degree program broadens access for women, first-generation to college, and low-income students, but grade disparities remain. *PloS one*, 15(12):e0243916. https://doi.org/10.1371/journal.pone.0243916.

NASA Exoplanet Archive (2022). https://exoplanetarchive.ipac.caltech.edu/ [Accessed: June 8, 2022].

National Research Council (1996). *National science education standards*. National Academies Press.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research, and Evaluation*, 20(1):2. https://doi.org/10.7275/hb2g-m060.

Perera, V., Mead, C., Buxner, S., Lopatto, D., Horodyskyj, L., Semken, S., and Anbar, A. D. (2017). Students in fully online programs report more positive attitudes toward science than students in traditional, in-person programs. *CBE—Life Sciences Education*, 16(4):ar60. https://doi.org/10.1187/cbe.16-11-0316.

Posner, G. J., Strike, K. A., Hewson, P. W., and Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2):211–227. https://doi.org/10.1002/sce.3730660207.

Prather, E. E., Slater, T. F., Adams, J. P., Bailey, J. M., Jones, L. V., and Dostal, J. A. (2004). Research on a lecture-tutorial approach to teaching introductory astronomy for non–science majors. *Astronomy Education Review*, 3(2). https://doi.org/10.3847/AER2004019.

Price, C. A. and Lee, H.-S. (2013). Changes in participants' scientific attitudes and epistemological beliefs during an astronomical citizen science project. *Journal of Research in Science Teaching*, 50(7):773–801. https://doi.org/10.1002/tea.21090.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., Schawinski, K., Szalay, A. S., and Vandenberg, J. (2013). Galaxy zoo: Motivations of citizen scientists. *Astronomy Education Review*, 12(1):010106. https://doi.org/10.3847/AER2011021.

Rudolph, A. L., Prather, E. E., Brissenden, G., Consiglio, D., and Gonzaga, V. (2010). A national study assessing the teaching and learning of introductory astronomy part ii: The connection between student demographics and learning. *Astronomy Education Review*, 9(1):010107. https://doi.org/10.3847/AER0009068.

Simon, M. N., Prather, E. E., Buxner, S. R., and Impey, C. D. (2019). The development and validation of the planet formation concept inventory. *International Journal of Science Education*, 41(17):2448–2464. https://doi.org/10.1080/09500693.2019.1685140.

Tien, L. T., Teichert, M. A., and Rickey, D. (2007). Effectiveness of a more laboratory module in prompting students to revise their molecular-level ideas about solutions. *Journal of Chemical Education*, 84(1):175. https://doi.org/10.1021/ed084p175.

Toomey, A., Strehlau-Howay, L., Manzolillo, B., and Thomas, C. (2020). The place-making potential of citizen science: Creating social-ecological connections in an urbanized world. *Landscape and Urban Planning*, 200:103824. https://doi.org/10.1016/j.landurbplan.2020.103824.

Volkwyn, T. S., Airey, J., Gregorcic, B., and Heijkenskjöld, F. (2019). Transduction and science learning: Multimodality in the physics laboratory. *Designs for Learning*, 11(1):16–29. https://doi.org/10.16993/dfl.118.

Volkwyn, T. S., Airey, J., Gregorcic, B., and Linder, C. (2020). Developing representational competence: linking real-world motion to physics concepts through graphs. *Learning: Research and Practice*, 6(1):88–107. https://doi.org/10.1080/23735082.2020.1750670.

Wallace, C. S., Chambers, T. G., Prather, E. E., and Brissenden, G. (2016). Using graphical and pictorial representations to teach introductory astronomy students about the detection of extrasolar planets via gravitational microlensing. *American Journal of Physics*, 84(5):335–343. https://doi.org/10.1119/1.4943035.

Wallace, C. S., Prather, E. E., and Duncan, D. K. (2012). A study of general education astronomy students' understandings of cosmology. part v. the effects of a new suite of cosmology lecture-tutorials on students' conceptual knowledge. *International Journal of Science Education*, 34(9):1297–1314. https://doi.org/10.1080/09500693.2012.677960.

Wallace, C. S., Prather, E. E., and Mendelsohn, B. M. (2013). Astro 101 students' perceptions of science: Results from the thinking about science survey instrument. *Astronomy Education Review*, 12(1). https://doi.org/10.3847/AER2012042.

9 Appendix A

Scale validity was assessed with both an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). The matched student responses were randomly split in half to create separate, independent data sets for the EFA and CFA, which left 162 responses for the CFA and 163 for the EFA. Pre- and post-survey responses were combined. There is no missing data in this analysis as only complete surveys were retained, and as it is Likert data, outliers are not a concern. We used the following thresholds to assess model fit: RMSEA $\leq$ 0.06, 90% CI $\leq$ 0.06, SRMR $\leq$ 0.08, CFI $\geq$ 0.95, and TLI $\geq$ 0.95 (Hu and Bentler, 1999; Brown, 2015).

A preliminary Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy was performed on all data subsets, with the measure of sampling adequacy (MSA) for both overall and individual items being $\geq$ 0.6. Normality was assessed with both Mardia's multivariate kurtosis test and Mardia's multivariate skewness test, with both indicating non-normality (p < 0.01 in both cases).

9.1 Exploratory Factor Analysis (EFA)

An EFA was performed using the principal axis factoring method with an *oblimin* rotation. This oblique rotation was used because all factors are based on self-reported comfort with scientific topics and it is unlikely that they are fully independent, so an oblique rotation is most appropriate (Costello and Osborne, 2005). In addition, oblique rotations perform comparably to orthogonal rotations in situations where the factors are truly independent making it the most appropriate choice given the uncertainty in our previously unvalidated instrument (Osborne, 2015). In order to determine how many factors to retain for the EFA, we performed a parallel analysis, visually inspected the scree plot, and considered our three-part curricular design model. Both the parallel analysis and the scree plot suggest that there are 3 factors. This is consistent with our theory as this instrument was adapted from a 4-factor evaluation by dropping all items related to a content-specific factor.

Given these results, we performed an EFA on two, three, and four factor solutions (M1, M2, and M3 respectively). The four-factor solution (M3) failed to converge and was discarded. The two-factor solution (M1) failed to meet critical thresholds for model fit (TLI = 0.781, RMSEA = 0.167, 90% CI = 0.046). The three-factor solution (M2) was the best fit (TLI = 1.016, RMSEA = 0, 90% CI = 0.32). In addition, communalities for each item in the three-factor solution were above 0.3, indicating that these items are appropriately grouped (Table 5).

9.2 Confirmatory Factor Analysis (CFA)

Model parameters for the CFA were obtained using a robust ML estimator (MLM) which includes a Satorra-Bentler correction factor to account for non-normalities. Based on the results of our EFA (as well as our pre-existing theory) we fit the CFA using

Table 5. Survey item numbers (see Table 2 for text), factor loadings, communality scores (h2), and uniqueness scores (u2) of three-factor EFA using an oblimin rotation and principal axis factoring method. A priori item/factor combinations are shaded in grey for convenience

Item	Data Literacy Self-Efficacy	Science Engagement	Citizen Science	h2	u2
4	0.07	0.71	0.00	0.57	0.43
5	-0.02	0.90	0.01	0.79	0.21
6	0.53	-0.02	0.22	0.45	0.55
7	0.87	-0.02	-0.03	0.71	0.29
8	0.59	0.12	-0.01	0.44	0.56
9	0.06	0.25	0.54	0.53	0.47
10	0.00	-0.02	0.99	0.95	0.046

three factors. This model fits acceptably well (RMSEA = 0.059, 90% CI = 0.028, SRMR =0.034, CFI =0.98, and TLI =0.96).

9.3 Limitations

The goal of this work overall was to develop an innovative curricular model. The assessment portion of this work was secondary to the curriculum development, and the results reported in this paper are preliminary. To increase the reliability of the factors, it would be best to have a minimum of three items per factor. Despite this limitation, the results of both our CFA and EFA had an acceptable fit. Further, our three-factor model with two items in two of the factors is still more reliable than performing univariate tests on each of the seven pre-/post- items.