Higher Education Quality Council of Ontario

An agency of the Government of Ontario

# The Role of Planetariums in Promoting Engagement and Learning

Michael Reid, Michael Williams, John Percy, Darren Hoeg, Kelly Lepo, Joanne Nazir and Gregory Paciga, University of Toronto

Published by

## The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

**Cite this publication in the following format:**

Reid, M., Williams, M., Percy, J., Hoeg, D., Lepo, K., Nazir, J., & Paciga, G. (2014). *The Role of New Planetariums in Promoting Engagement and Learning*. Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

# Table of Contents



A separate appendix is available in English only from heqco.ca.

# List of Figures



# List of Tables

# Executive Summary

Most astronomers teaching undergraduate astronomy aspire to connect their students directly with the night sky. In the same way that a biologist might want her students to actually handle live specimens or a geologist for his students to chip away at real rocks, astronomers want their students to actually see and observe planets, stars and galaxies. Sadly, the combination of urban light pollution, unpredictable weather and daytime teaching schedules make this impractical. This is especially the case for high-enrolment survey courses, which present the additional complication of huge numbers of students to schedule.

An increasingly common strategy is to teach astronomy in digital planetariums: domed rooms on whose ceilings can be projected fantastically detailed representations of the night sky. Planetariums are, in many ways, more useful than the actual sky: they can be used during the day, are not subject to changeable weather, and can be manipulated to show sights not normally visible in the actual sky. Even better, digital planetariums can have control interfaces which are simple enough that almost anyone can use them – ours uses an off-the-shelf video game controller.

Despite their growing popularity, there is relatively little literature on the effectiveness of digital planetariums as teaching tools. We sought to determine whether a small digital planetarium (ours has a capacity of 25 people) enhances student learning and engagement in a large first-year survey course in astronomy for non-science students. In particular, we wanted to know whether giving students direct control of the planetarium experience improved their conceptual gains and their engagement with the subject matter.

We studied the use of planetarium shows in a 1,350-student introductory astronomy course at the University of Toronto. Each student was assigned three experiences: an interactive planetarium experience, a non-interactive planetarium show conducted by a teaching assistant, and a discussion-based tutorial outside the planetarium. The students' conceptual gains were assessed using multiple-choice tests and their impressions of their experiences were gathered using focus group discussions.

Our findings demonstrate very clearly that a digital planetarium is not a magic bullet for teaching astronomy. There is a definite "wow factor" to the planetarium experience which improves students' level of engagement, but the planetarium does not automatically increase students' conceptual gains. More specifically, we found no overall significant statistical difference between the conceptual gains of students who experienced interactive planetarium shows, non-interactive planetarium shows, and discussion-based tutorials outside the planetarium. All three experiences were approximately equally good at increasing student conceptual gains on multiple-choice tests. Depending on the topic being covered, sometimes discussion-based tutorials were the best method.

This study is preliminary – a first attempt at assessing the best way to use a digital planetarium as a teaching tool. One of our main conclusions is that first-year non-science students are not well equipped to manage their own learning in a planetarium environment without a lot of scaffolding. The "wow factor" of being in a planetarium does not necessarily draw them in quickly or deeply enough to overcome the technical difficulties of operating the planetarium. Moreover, the students were very clear in expressing that, though they saw the planetarium as a potentially very useful learning tool, they would need to spend a lot of time in it getting used to the controls before useful learning could result.

We looked for statistically significant effects associated with the size of the student group involved in each of our three interventions, the teaching assistant overseeing the experience, and the gender of the students. We found no significant correlations with group size or teaching assistant identity, but did find some interesting

trends associated with gender. Female students scored significantly lower on the pre-tests than male students and generally produced lower conceptual gains in all three intervention types than male students. These gender-based differences were not always statistically significant but they persisted throughout the entire data set.

We recommend significant additional study of the use of digital planetariums as teaching tools. Our results point to several ways to optimize the planetarium experience: spend more time teaching the students how to use the planetarium, scaffold the self-directed experience more rigorously, and spend more time overall in the planetarium. In addition, we recommend further investigation into the possible gender differential in both students' incoming conceptual understanding of astronomy and their conceptual gains as a result of all three of our intervention types.

# Introduction

Most universities offer some type of introductory astronomy course that is open to non-science majors. These courses are often very popular and have enrolments in the hundreds or even thousands. They present unique challenges because most students enter university having had little formal exposure to astronomy, and almost none of it from astronomy specialists. Some of them have not taken a science or math course in three or four years. Having heard about recent discoveries of Earth-like planets, the search for dark matter and the study of the origins of the universe, they arrive on university campuses eager to learn more. Yet they are often poorly prepared to begin the study of astronomy.

Astronomy is a relatively small discipline. The entire population of astronomy majors at even a large university can easily be exceeded by the enrolment in a single semester of an introductory astronomy course (generically ‘AST 101’ hereafter). As such, large-format introductory astronomy courses can actually constitute the main teaching responsibility of astronomy departments, measured by sheer number of students served.

Despite the popularity of these courses, their centrality to the teaching of astronomy and their value as part of a well-rounded education, little is known about how to teach them well. They tend not to receive as much attention as classes dedicated to students in physical science majors, such as introductory physics or chemistry. Astronomers looking for evidence-based approaches to teaching are heavily reliant on the physics education literature which, although an excellent resource, does not address the particular needs of astronomy learners, particularly those science-curious but math-phobic learners who typify first-year non-science majors taking AST 101.[1]

Teaching astronomy is not quite like teaching other sciences. Unlike physics, chemistry and biology, the subject matter in astronomy cannot be touched or experimented on directly. Thus, teaching techniques which rely on direct experimentation must be adjusted for use in astronomy. For the most part, astronomy is a science of taking pictures of the night sky. Given that astronomy is almost always taught during the day, this poses significant challenges to any instructor who wants to involve her students in the acquisition of astronomical data. The situation is analogous to trying to teach introductory chemistry, but only between the hours of one and five in the morning and without letting the students touch the chemicals. Even when astronomy classes can be taught in the evening, most of them are taught in cities large enough that light pollution is a serious problem, greatly impeding attempts to view the sky. At rural universities, where light pollution is less of a problem, unpredictable and often chilly weather makes it very difficult to schedule regular sky-viewing sessions.

Given the difficulties of bringing AST 101 students in contact with the sky, astronomers increasingly turn to planetariums to provide simulated views of the sky. The use of planetariums in teaching has yet to receive the same level of scrutiny as other teaching interventions. This is partly because affordable, highly capable digital planetariums are a relatively recent innovation, becoming available only in the last ten years or so. Prior to that, most Canadian universities could not afford a planetarium. Those that could often had the old style of planetarium, which could only display the sky as seen from Earth and in very limited detail. This older style of planetarium emphasizes concepts which are of low priority in teaching astronomy, such as the arrangement

[1] Until recently, astronomers seeking information about how best to teach their subject could rely on but one journal, Astronomy Education Review (AER). Sadly, AER has recently announced plans to cease publishing at the end of 2013, leaving the astronomy community without a focused education research journal.

of the constellations, the rise and set times of stars, and the paths followed by the planets. These concepts rarely form the core curriculum of modern astronomy courses. To continue the chemistry analogy, using an older-style planetarium to teach modern astronomy is like trying to teach advanced biochemistry using only water and salt.

Modern digital planetariums are far more capable than their older counterparts, at a fraction of the cost. They share in common only their basic physical setup: a domed room on whose ceiling some portion of the sky can be projected. But modern systems, including the one used in our study, offer vastly more opportunities for learning. The digital planetarium at the University of Toronto (like most others of its kind) can render in real time any part of the observable universe. It can and is modified often to incorporate recent findings, such as the discovery of new planets and galaxies. Using such a system, students can explore the cosmos in a way not possible with any other technique. For example, students can:

- Fly around the solar system, viewing it from any angle as time passes at an accelerated rate, allowing students to relate the shape of the solar system and the motions of the bodies in it to theories of its formation.
- View all currently known extrasolar planets (planets in other solar systems) on a three-dimensional model of our Milky Way galaxy, allowing students to assess the distribution and proximity of these systems and how these factors relate to our detection methods.
- Compare views of the night sky as seen from the surfaces of any bodies they choose, to help develop their sense of place.
- Quickly recognize that the stars making up any given constellation are not physically associated with one another, by zooming outward from the Earth and watching the lines connecting the stars distort as the viewer passes through them.
- Gain an appreciation for the scale of the universe by flying through it at or above light speed.
- Rapidly cycle through all of the relevant spatial scales in the cosmos (planets, stars, galaxies, galaxy clusters, etc.), seeing how they are related to one another.

This ability to change scale and position rapidly in both space and time was not present in the older style of planetarium. Thus, studies of the use of older styles of planetariums have limited applicability when teaching these concepts using modern digital planetariums.

Another appealing feature of digital planetariums is that they can be controlled using simple, programmable controllers. Gone are the complex consoles covered with buttons, dials and switches. The University of Toronto's planetarium can be controlled using only a standard video game controller and an optional touch-screen computer monitor.

The newer style of digital planetarium is more interactive and immersive, so one might wonder whether it is also more engaging for students and, separately, whether this engagement promotes learning. In older planetariums, the student is usually restricted to being a passive member of the audience and can only see the universe from a very limited perspective. Generations of students have been inspired by shows conducted in this style of planetarium, but we are interested in discovering the effects of the newer, more interactive style of planetarium, particularly on the novice learner of astronomy. In newer digital planetariums, students can lead shows themselves with relative ease. We would like to know whether this ability to explore in a more self-directed way can promote greater engagement with the subject matter and perhaps deeper learning.

The simulation of the universe contained within digital planetariums (including ours) is so complete that students can explore almost any corner of the universe about which they might be curious. The experience of being in a digital planetarium, of moving around through space and time at speeds not physically possible, often evokes expressions of awe and wonder from students who have never thought of the universe in this way before and certainly never imagined controlling it themselves. But can novice learners of astronomy harness the potential of such a system when allowed to control it themselves? Will it engage them and promote learning?

Our goal in this study was to evaluate the use of digital planetariums for two main purposes: to help students learn about difficult astronomical concepts and to help increase their level of engagement with astronomy.

## Research Questions

Our study seeks to answer three specific questions:

1. Do interactive planetarium shows help students learn about concepts in astronomy, particularly those based on temporal and/or spatial reasoning?

2. Are interactive planetarium shows better at increasing students' understanding of astrophysical concepts than non-interactive planetarium shows?

3. Does our highly engaged mode of teaching, including such elements as in-class clicker quizzes, online interaction and interactive planetarium shows, help improve student attitudes toward and engagement with astronomy?

The first of these questions is intended to target the supposed strengths of a digital planetarium system. That is, if students are able to manipulate digital models of astrophysical systems rather than just looking at static images or pre-rendered videos of them, does this help them understand those systems better? For example, does hands-on experience with a digital planetarium help students grasp the size of the solar system better than seeing static images? Does it help them understand the orientations of planetary orbits relative to one another better than seeing static images of the solar system from multiple perspectives?

The second question is intended partly as a control for the first, to separate the effect of being in the planetarium from being *in control* of the planetarium. It is possible that students might learn more effectively when they are in charge of their own learning. However, it is also possible that even our simplified planetarium is still sufficiently complex that strong guidance from an expert is necessary to make progress. We emphasize, however, that none of our pedagogical interventions would be considered "inquiry-based" in a rigorous sense, in that we did not provide the detailed scaffolding and time for tinkering that characterize true "inquiry." We only sought to test whether, in the specific case of using a planetarium to learn astronomy, the inclusion of a basic level of interactivity helps or hinders student learning.

The third question is intended to examine the effectiveness of planetarium shows in increasing student engagement with astronomy, as part of an overall "highly engaged" style of teaching. By "highly engaged", we mean that students in our AST 101 course have many different ways to engage with each other, with teaching assistants (TAs) and with the instructors. They are also offered several different extracurricular activities, some of which are optional. We wanted to know whether, in this context, planetarium shows of the type we are able to offer are playing a useful role in engaging students.

We emphasize that we have not attempted to answer these questions – particularly the first two – in an idealized laboratory setting. This study does not address the question of whether it is *possible* for interactive planetarium shows to produce higher conceptual gains than other types of shows. Instead, we are trying to compare the effectiveness of interactive and non-interactive planetarium shows, *as deployed in a large course, using a typical group of teaching assistants who have various backgrounds and levels of competence as instructors.*

It is possible that, working with small groups of students under laboratory conditions, one could arrive iteratively at the best possible method for using a planetarium to increase student conceptual gains and engagement. This would entail the development of a concept inventory tailored to concepts that can be taught in a planetarium. It would also entail many repetitions of the experiment under different conditions. That was not the purpose of this study. Instead, we wanted to study the use of the planetarium in situ to see whether, in a most plausible use case, the level of interactivity in a planetarium show has a meaningful impact on student outcomes.

# Previous Research on the Use of Planetariums in Teaching

It is well established that many students of all ages struggle with learning science (e.g., Hodson, 2009). Science educators have suggested several factors, including sociocultural contexts, politics and gender (Aikenhead & Michell, 2011; Baker, 2002; Roth & Calabrese Barton, 2004), as explanations for this situation. While many of these explanations are controversial, one point upon which many agree is that science concepts that require abstract thinking and complex visualizations like those found in physics and astronomy are particularly difficult for students to grasp (Prather et al., 2009; Taylor et al., 2003). This issue becomes more of a concern as students reach the postsecondary level, where science concepts become increasingly abstract, particularly in astronomy. Computer simulations and other interactive technologies are increasingly being proposed as powerful techniques to help students learn complex scientific concepts (Edelson, 2001; Winberg & Berg, 2007; Songer, 2007).

Much of the existing literature on the use of planetariums in education concerns the older style of planetarium and is focused on elementary-level students. Take, for example, the cause of the seasons. This concept is a perennially difficult one for students and teachers alike to understand, in part because it involves spatial and temporal reasoning. Atwood and Atwood (1996) found that, of 49 preservice elementary school teachers, only 1 was able to provide a correct explanation of the cause of the seasons; the majority of the remainder provided an incorrect alternative explanation. Plummer (2008) found that young children who learned about the seasons using a planetarium-based approach showed significant improvement over those using more traditional approaches. One of our goals is to determine whether students who are able to manipulate a three-dimensional model of the solar system are better able to understand concepts related to the size and shape of the solar system and to apply that understanding in novel contexts.

Mallon and Bruce (1982) found that participatory planetarium shows produced significant impacts on both student understanding of and appreciation for science. The subjects in these studies were young children, whereas we aim to test the effects of interactive planetarium shows on the attitudes of undergraduates. Also,

these studies considered verbal interaction to be "participation," whereas we propose to allow the students to manipulate the planetarium projection itself using a simple controller. Contrary to Mallon and Bruce (1982), a meta-analysis by Brazell and Espinoza (2009) of 19 studies showed that, while planetariums are effective tools for enhancing student learning, they have not yet proven to be an effective tool for enhancing student attitudes toward science. Brazell and Espinoza (2009) also pointed out that the research methods employed in many of the studies they cite lacked rigor, suffered from small-N statistical effects, and generally did not concern the interactive use of a planetarium. We hope to address all three of these issues in this study.
The literature on student engagement is considerably more robust than that on the use of planetariums in teaching. The general consensus is that, when teaching physical sciences, the "interactive engagement" style of teaching produces greater conceptual gains, as long as the engaging activities are implemented in a pedagogically rigorous way. Extensive work by Richard Hake and his collaborators, summarized in Hake (2002), clearly demonstrates that a variety of interactive teaching methods, such as hands-on activities and peer instruction, reliably produce significantly higher conceptual gains in introductory physics classes than traditional teaching methods (i.e., uninterrupted lecturing). After investigating numerous other possible explanations for these findings, Hake concludes that the effect really is caused by the higher level of student engagement in the learning process. Thus, one of our goals is to determine whether or not planetarium-based pedagogy is engaging, separately from its effect on conceptual gains.

# Methods and Methodology

This study was carried out in the course "AST 101: The Sun and Its Neighbours" during the fall semester of 2012 at the University of Toronto. AST 101 began with 1,350 students; of these, 1,201 students ultimately completed the course. The course is open only to those not majoring in the physical sciences (i.e., it excludes students majoring in physics, chemistry, biology, computer science, mathematics and engineering), but admits some students in the life sciences stream. The majority of the students in the class – and therefore the majority of those participating in the study – is drawn from across the humanities and social sciences, including students studying fields as diverse as criminology, commerce, music, physical education, drama and foreign languages. The student population at the University of Toronto is very diverse and, as a result, the study participants comprise a highly varied set of ethnicities, first languages, national origins and educational backgrounds.

To help interpret the results of our experiment, it will help to know a little about the planetarium itself. Our planetarium is an inflatable GeoDome planetarium, purchased in 2009 from The Elumenati.[2] It seats 25 students comfortably. It uses the Uniview[3] software package to display any part of the observable universe currently known to science. A picture of the dome itself is shown in Figure 1.

[2] http://www.elumenati.com/products/geodome/
[3] http://sciss.se/uniview.php

**Figure 1: A photograph of our GeoDome planetarium, positioned in a stadium for a special event. Normally the planetarium is situated in the basement of our astronomy building**

The study was carried out using a mixed methodology (Tashakkori & Teddlie, 1998). It consisted of several independent quantitative and qualitative methods of data collection. Analysis involved both qualitative and quantitative procedures and sometimes transformation of data from one form to the other to reveal trends and patterns salient to answering the research questions. According to Creswell and Plano Clark (2007), this type of methodology is particularly suitable for conducting studies on topics about which little is established and which require in-depth exploration. Early on, the research team identified this as the context that best described our topic of study.

We collected two main forms of data:

1. Pre- and post-tests for every student in every planetarium show and tutorial that was part of the experiment (primarily to help answer research questions 1 and 2 above).

2. Focus group interviews with a small subset (approximately 40) of study participants (primarily to help answer research question 3 above).

We kept track of limited demographic data for each student, including gender, year of study and tutorial membership.

We recruited students for this study using the course Learning Management System (LMS). Our LMS is Blackboard, referred to at the University of Toronto and throughout this document as "Portal." When students logged into the Portal site for this course, they were presented with an invitation to participate in our study. The consent and research information forms are included in Appendix A and Appendix B, respectively. The total number of students who initially consented to participate in the study was 1,005 (of whom 440 were men and 565 were women).

The organization of such a large course is rather complicated, with many tutorial sections and teaching assistants. Every week, the students were expected to attend two lecture hours and one tutorial hour. Our experiment was run during the tutorials. We ran three complete iterations of our experiment over the course of the semester. During each iteration, the students were broken into three intervention groups:

1. The “T” group, which experienced a conventional tutorial
2. The “Pt” group, which experienced a planetarium show led by a TA
3. The “Ps” group, which experienced a planetarium show led by the students themselves and their peers

Over the course of the three iterations of the experiment, each tutorial section (and therefore every student) cycled through all three groups. This ensured that each student had a similar experience to the others and that none of them benefitted or suffered more than others from our different intervention types. This approach was endorsed as part of our research ethics review. Note, however, that the subject matter of the planetarium show varied from one iteration of the experiment to another, in keeping with the progression of the curriculum in the course.

To understand the results of our experiment, it is important to understand what went on during each intervention type. In all three intervention types, the starting and concluding activities were the same. Each session began with a pre-test consisting of three questions administered using an audience response system (iClickers). Each session concluded with a post-test administered in the same way. Each intervention type is described more fully below:

- **The “T” or “Tutorial” Intervention**

  Students in this group experienced a “standard” tutorial, typically lasting 50 minutes but later reduced to 30 minutes, as will be described below. During a tutorial, the TA would guide the students through a discussion of the curricular content for that week of the course. For each tutorial, the TAs were given a small number of prepared slides, largely consisting of astronomical images, a discussion plan, and often a set of small-group activities. With 12 TAs of varying experience levels delivering tutorials across 32 tutorial sections, the consistency of the implementation of the tutorial plan varied considerably. Some TAs – particularly the less experienced ones – tended to adopt a minimally interactive “lecture-style” approach to leading tutorials. Other TAs adopted a more interactive approach, spending most of the time asking and answering questions. We emphasize consistency in the delivery of the tutorial plans, but it is not always possible to achieve consistency in practice.

- **The “Pt” or “Planetarium: TA-led” Intervention**

  Students in this intervention experienced a planetarium show conducted by a trained TA. The show is presented “live,” not pre-recorded, but the TA was directed to follow a plan provided for the show. In this way, the T and Pt groups are similar: in both cases, the TA is “in charge” of the learning process. In both the T and Pt intervention types, the TA is free to interact with the students, answer their questions, and even let the students shape the experience. For example, the students might ask to see some particular sight in the planetarium or might raise a particular topic of discussion in the tutorial. The extent to which a given TA provided that higher level of interactivity depended strongly on the TA. The group sizes for this intervention type ranged from 1 to 20 students, with a mean of about 15.

- **The "Ps" or "Planetarium: Student-led" Intervention**

  Students in this intervention were given total control of the planetarium. They were provided with a simplified control interface, in the form of a wireless video game controller whose buttons had been programmed to control certain aspects of the planetarium relevant to the concepts they were asked to investigate. For example, in an activity concerning the shape and size of the solar system, one joystick might zoom the field of view, while another rotated it. The buttons could turn on the orbits of planets in our solar system, toggle grids showing the physical scale of the phenomena being shown, and so on. The students in this group were provided with "challenge questions" that they were asked to answer during the time provided. As in the Pt shows, an experienced TA was present to help out as needed, but this person was instructed not to direct the students' exploration. The TA was also asked not to touch the controls, except to show the students how they worked. The goal was for the students to use a simplified version of a complicated system to make discoveries on their own, motivated by the challenge questions and the knowledge that there would be a post-test related to those questions.

  Note that this intervention type, while student-led, was not a fully "inquiry-based" experience. In the 30 minutes available for these shows, a fully inquiry-based activity was not possible. Nevertheless, we wanted to test to see whether giving control of the show to the students during the small amount of time available was better than having a TA guide them through the entire show.

  We tried to keep the groups as small as possible for this intervention type, but numerous factors beyond our control, such as the amount of planetarium time available and the students' schedules, constrained the group sizes. Typical group sizes ranged from 5 to 10 students.

We ran three iterations of the experiment. Initially, we intended there to be no differences between the iterations of the experiment, so that they would all be completely inter-comparable. In practice, we found it necessary to modify the protocol slightly after the first iteration of the experiment. This should be kept in mind when interpreting the results: iterations 2 and 3 are comparable to one another, but they were slightly different from iteration 1. The differences between iteration 1 and iterations 2 and 3 were:

- In iteration 1, the T group had 50 minutes of instruction between the pre- and post-tests. This was reduced to 30 minutes in iterations 2 and 3 to match the duration of the Pt and Ps interventions.

- In iteration 1, we observed that some of the TAs were getting too involved in the Ps intervention, doing too much of the work for the students. To address this, we changed the style of the plan given to TAs for the Ps shows in interventions 2 and 3. Instead of giving them a script for the activity, they were given a set of goals to lay out for the students and prompts they could use to help guide the students.

## Pre- and Post-Tests

A significant challenge in evaluating the students' conceptual gains for this project was the lack of available validated concept inventories. By "concept inventory," we mean a standardized multiple-choice diagnostic test of students' existing knowledge of a topic. The canonical concept inventory in the physical sciences is the Force Concept Inventory (Halloun & Hestenes, 1985), which assesses students' understanding of Newton's Laws. There are a few validated concept inventories for astronomy, such as the Light and Spectroscopy Inventory (Bardar, 2006) and the Astronomy Diagnostic Test (Deming, 2002). However, none of the available tests were well matched to our pedagogical goals for these interventions (they are generally geared toward more science-intensive courses and cover a wider range of topics than our interventions did). Also, the compressed time available for the intervention did not allow the administration of a long test. Instead, we came up with our own pre- and post-tests, based on our long experience of teaching this course and our particular pedagogical goals. We did this by drafting a long list of questions that addressed our pedagogical goals for the intervention, then circulating them among expert colleagues for commentary and refinement. We narrowed the list until we had six questions for each intervention: three "pre" questions and three "post" questions (a total of 18 questions across all three iterations of the experiment). For each intervention, the pre and post questions were matched in style and content but not identical. Each pair tested the same concept or concepts but with different wording. While we recognize that this introduces some difficulties in comparing the results from the pre-test to those from the post-test, we thought that it was important that the pre- and post-tests not be identical. Given that the pre- and post-tests were administered within 30 minutes of one another, a student's performance on identical questions could have improved simply through repetition and rehearsal rather than through genuine conceptual gains. This is why we decided to vary the wording of the questions. A sample of the pre- and post-test questions is included in Appendix C.

All students in the course were assigned to participate in all three intervention types, whether or not they had consented to be part of the experiment. Only the scores of consenting students were included in this study. Indeed, this was the only difference between consenting and non-consenting students.

## Focus Groups

When students gave their consent to participate in this study, they were asked to indicate whether they would consent to be asked to participate in focus groups. A total of 605 students consented to be invited. Our goal was to have five focus group meetings with ten students per meeting. Invitations were sent to participants by e-mail. All participants were given a $20 gift card for participating (to their choice of either iTunes or Starbucks). Invitees were given 48 hours to accept or reject the invitation. In the end, 280 invitations had to be sent to fill the 50 available spots, though not all of the 50 confirmed registrants actually came to their focus groups.

Each focus group was administered by a pair of research assistants. The same research assistant facilitated all of the groups, asking a pre-determined set of questions as well as improvised follow-up questions. The second research assistant monitored the audio recording equipment and kept notes on the order of the speakers (to assist in later transcription). All five focus groups were later transcribed and the transcripts checked by other members of our research team. The same research assistant who facilitated the focus groups also performed a thematic analysis of them, using the NVivo software package.

# Findings

## Student Demographics and Statistical Summary

To aid in the interpretation of our results, we present a demographic snapshot of the students who participated in the project. As mentioned earlier, of the 1,201 students who completed the course, 1,005 consented to participate in our study. Of these, 852 successfully completed both the pre-test and post-test in at least one iteration of the experiment. Of the study participants, 41% were men and 59% were women, with one student declining to specify a gender. This is approximately representative of the entire undergraduate student body at the university, which consists of 44% men, 56% women, and a small number of students with undisclosed genders.

In terms of year of study, 64% of the study participants were in first year, 15% in second year, 9% in third year, 11% in fourth year, and 0.004% (3 students) in year five or above.

The sizes of the student groups in each intervention type varied considerably. At the start of the semester, the mean number of students in each tutorial section was 42. By the end of the semester, this had fallen to a mean of 37.5 students. However, the number who showed up on any given day to participate in an intervention varied significantly. The mean number of study participants who took part in each intervention type was 7 for the Ps and Pt groups and 9 for the Tutorial groups. The standard deviation on all three numbers is 3. Given the consent rate for this experiment, these numbers of consenting students should be, on average, about 20% lower than the total number of students in the room during each intervention (i.e., including those who did not consent for their participation to be tracked).

## Quantitative Findings

To make a quantitative determination of the relative gains produced by our various intervention types, we examined the normalized gains produced by each one. By normalized gains, we mean the fractional improvement in a student's performance on the post-test, normalized to her total potential score. We use the prescription articulated by Bao (2006):

$$g(x,y) = \begin{cases} \dfrac{y-x}{1-x} > 0 \ (y \geq x) \\ \dfrac{y-x}{x} < 0 \ (y < x) \end{cases} \tag{1}$$

where x is the student's fractional pre-test score and y is her fractional post-test score. The values of g(x,y) at the singular points are defined by:

$$g(x,y) = \begin{cases} 1 \ (y = x = 1) \\ 0 \ (y = x = 0) \end{cases} \tag{2}$$

Thus, the gain is normalized: a gain of 1 means that the student improved by the maximum amount possible between pre- and post-tests, while a gain of 0.5 means the student made only half of the total possible improvement (e.g., going from a pre-test score of 50% to a post-test score of 75%, or 80% to 90%). Our quantitative results are organized according to the effects of: intervention type, TA, group size and gender. A discussion of the statistical significance of these results follows the presentation of the results themselves.

### Intervention Type

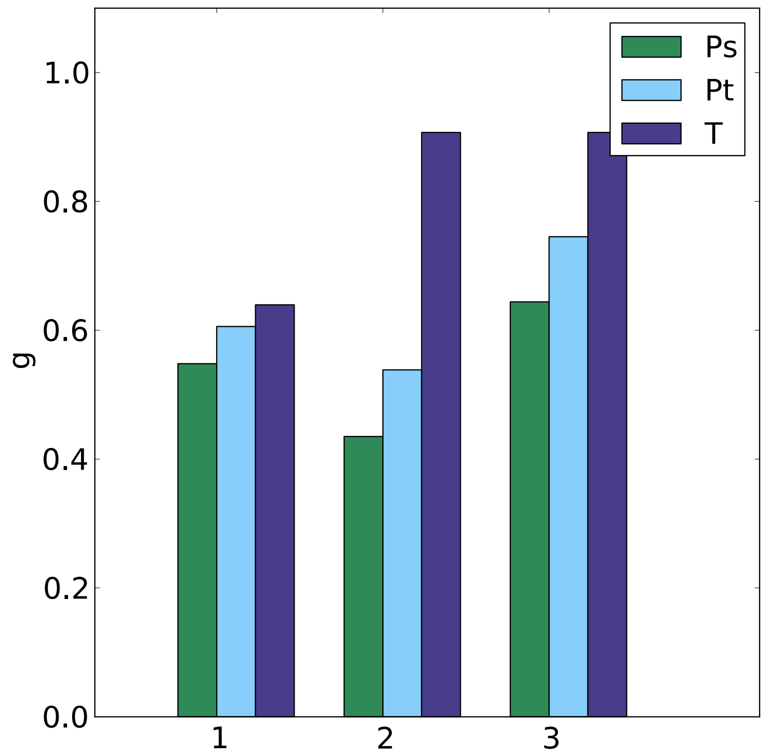
Figure 2 shows the normalized gains for each question separately for all three intervention types in the first iteration of the experiment. This figure shows that, for all three questions, the greatest conceptual gains were consistently achieved in the tutorial setting, not in the planetarium. This was the finding that prompted us to change our experimental technique for the second and third iterations of the experiment. In this first iteration, the tutorials were allowed to run for almost twice as long as the planetarium shows, so that students in the T groups were getting twice as much time to contemplate the relevant astronomical concepts as were the students in the Ps and Pt groups.[4] To control for this effect, we equalized the duration of the tutorial and planetarium experiences for the second and third iterations of the experiment.

Figure 3 and 4 show the same analysis for iterations 2 and 3 of the experiment. In these two iterations, with the times for all intervention types equalized, the apparent benefit of the tutorial intervention over the two planetarium types of intervention vanishes. The effect of the question itself becomes more apparent, too: it is clear that question 2 in iteration 2 of the experiment was much better suited to the tutorial format than to the student-led planetarium format. Interestingly, this question, which concerns the shape of the solar system, is precisely the sort of spatial-reasoning question which we expected might be best-suited to the planetarium (the question itself is shown in Appendix C). Further investigation of the suitability of digital planetariums for teaching specific concepts seems warranted.

None of these figures includes uncertainty intervals. Our pre- and post-tests each included only three questions so the resulting gains are relatively coarsely discretized. The coarse discretization of the gains means that they are not normally distributed (i.e., not Gaussian) and therefore the uncertainties on the gains cannot be properly represented with conventional uncertainty bars, showing the standard deviation of the data. Instead, to assess statistical significance, we have performed pairwise Kolmogorov-Smirnov (K-S) tests on our data. The K-S test compares two different data sets without reference to their parameterization and returns the probability that they are drawn from the same underlying (and unknown) parent distribution. Thus, we can measure the probability, p, that the distributions of answers in two experimental groups are drawn from the same parent distribution without having to assume that either set follows a Gaussian distribution. For our purposes, we set our threshold of statistical significance at $p \leq 0.05$. This means that, when comparing two sets of student answers (e.g., those from one intervention type to those from another), we interpret the difference to be significant if the likelihood of the two data sets being drawn from the same distribution is less than 5%.

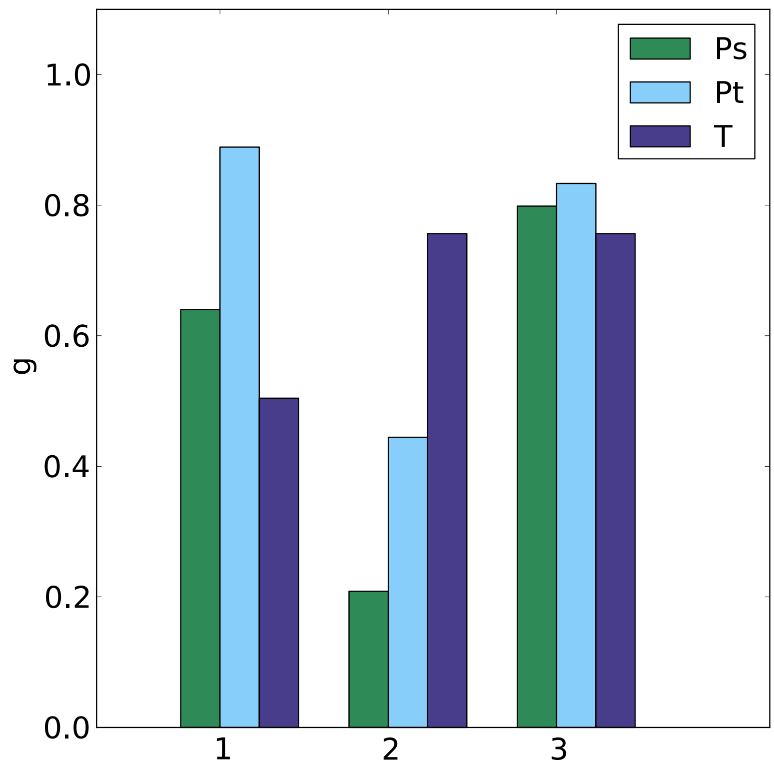
[4] The longer time allowed for tutorials was simply an oversight on our part. Traditionally, the tutorial sessions for this course have been 50 minutes long and the planetarium shows have been 25 minutes long (to allow for smaller planetarium groups). It did not occur to us to equalize the times until after the first iteration of the experiment.

**Figure 2: – Normalized Gain as a Function of Intervention Type for Iteration 1 of the Experiment**



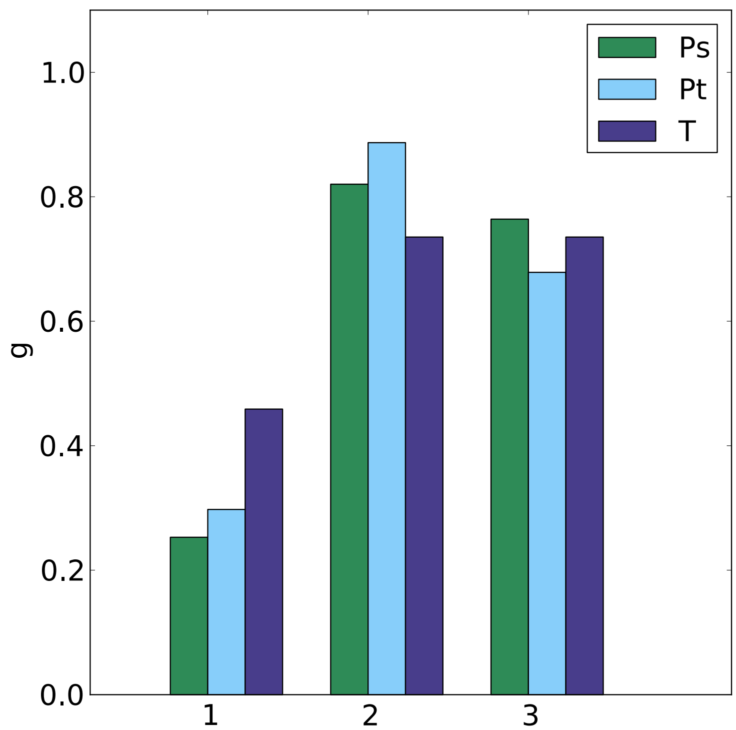

Each bar shows the normalized gain, g, between the pre- and post-tests for each of the three questions (1, 2, and 3). The results for the student-led and TA-led planetarium shows are shown in green and light blue, respectively. Those for the tutorials are shown in purple. See the text for an explanation of the statistical significance of these results. Note that, for this iteration of the experiment, the time allowed for tutorials was twice as long as that allowed for planetarium shows; in future iterations, the time allowed was equalized.

**Figure 3: Normalized Gain as a Function of Intervention Type for Iteration 2 of the Experiment**




Each bar shows the normalized gain, g, between the pre- and post-tests for each of the three questions (1, 2, and 3). The results for the student-led and TA-led planetarium shows are shown in green and light blue, respectively. Those for the tutorials are shown in purple. See the text for an explanation of the statistical significance of these results.

**Figure 4: Normalized Gain as a Function of Intervention Type for Iteration 3 of the Experiment**



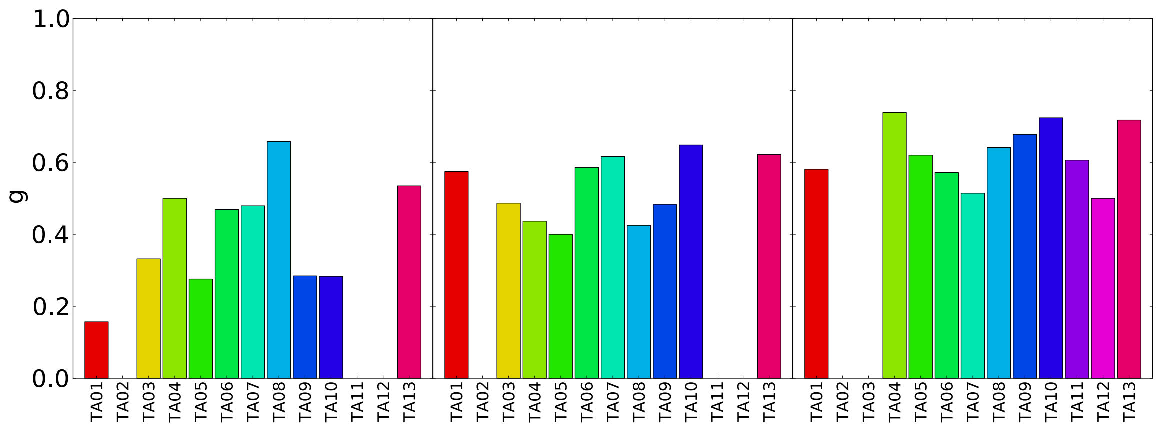

Each bar shows the normalized gain, g, between the pre- and post-tests for each of the three questions (1, 2, and 3). The results for the student-led and TA-led planetarium shows are shown in green and light blue, respectively. Those for the tutorials are shown in purple. See the text for an explanation of the statistical significance of these results.

### *TA Effects*

Perhaps the largest uncontrolled variable in our experiment was the skill level of the TA who led each group of students through each activity. We wanted to evaluate the effectiveness of our methods *in situ*, not under controlled conditions with hand-picked instructors. In our experience – and we expect that of most other instructors – the pool of TAs assigned to any given course is very heterogeneous. Our TAs have differing personalities, educational backgrounds, fluency in English, teaching styles and amounts of TA experience. Some are fresh entrants to graduate school, having had little undergraduate astronomy education themselves, while others are seasoned PhD candidates who might have been TAs for five or six years. Furthermore, at our institution, TAs do not receive extensive training for their teaching roles. They received only a few hours of union-mandated training, plus whatever else the individual instructors are able to provide. Given all of these factors, we wondered whether there would be marked differences in the student gains as a function of the identities of their TAs.

Figure 5 demonstrates the heterogeneity of our TA pool. While it does not show any strong trends, there are many interesting cases. Here, we interpret the variations in the normalized gains as partly reflective of the effectiveness of the TAs as instructors in different settings (although we recognize that other factors may have been involved). For example, TAs 1, 2 and 10 all produced substantially higher gains in the Pt intervention than in the Ps one. This suggests that the TAs could use help in delivering the Ps shows more effectively or that their skills are best suited to the Pt shows. TA 7 was approximately equally effective in all intervention types, always producing gains of about 0.5. The low gains achieved by TAs 1, 5, 9 and 10 in the Ps intervention type suggest that they might benefit from additional training on how to deliver Ps shows most effectively.

**Figure 5: Normalized Gains, g, produced by different TAs, shown for all three Intervention Types: Ps (left), Pt (centre) and T (right)**



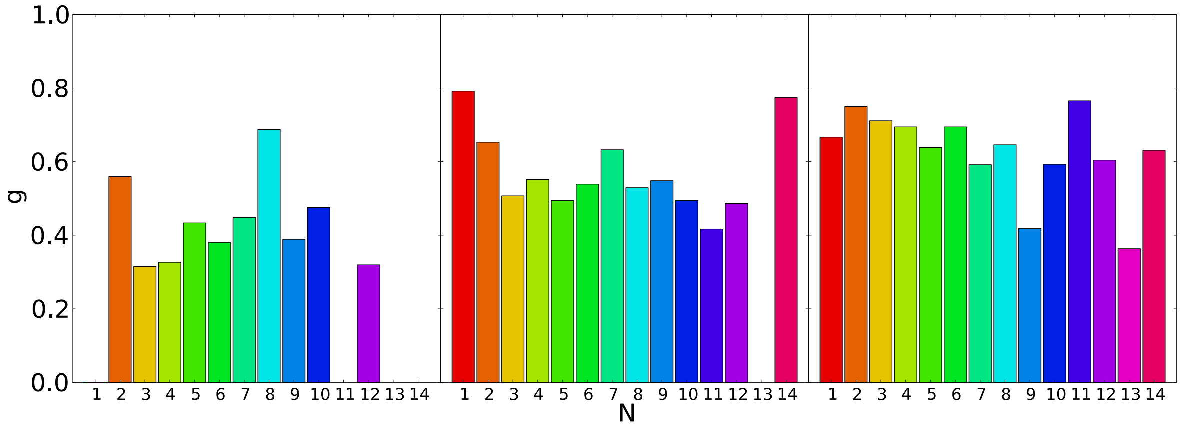

Each column represents a different TA. Not all TAs participated in all intervention types, so some columns are missing. The identities of the TAs remain the same from one intervention to the next, so that TA01 is comparable from one graph to another.

## *Group Size*

One might reasonably anticipate that the effectiveness of an intervention type would depend on the number of participating students – on "class size," or what we are calling "group size." Particularly in the Ps intervention group, where students are interacting with one another in the planetarium and sharing access to a limited resource (the planetarium controller), we would expect that smaller group size might yield higher gains. In general, we do not find this to be the case. Figure 6 shows the normalized gains for students participating in each intervention type, separated by group size.

While there were a few statistically significant differences in the distribution of normalized gains between groups of different sizes, there is no clear trend in normalized gain versus group size. It is not the case, for example, that smaller groups regularly produced higher gains. Also, the most significant differences among group sizes occur in TA-led planetarium shows and tutorials, where we did not expect the strongest group size effects. We had expected that group size would be a stronger determinant of success in the student-led planetarium shows, where the students are sharing a limited resource – the planetarium controller – than in the other intervention types. These data lead us to conclude that group size is not a significant determinant of the effectiveness of any of our intervention types.

**Figure 6: Effect of Group Size on Normalized Gains**



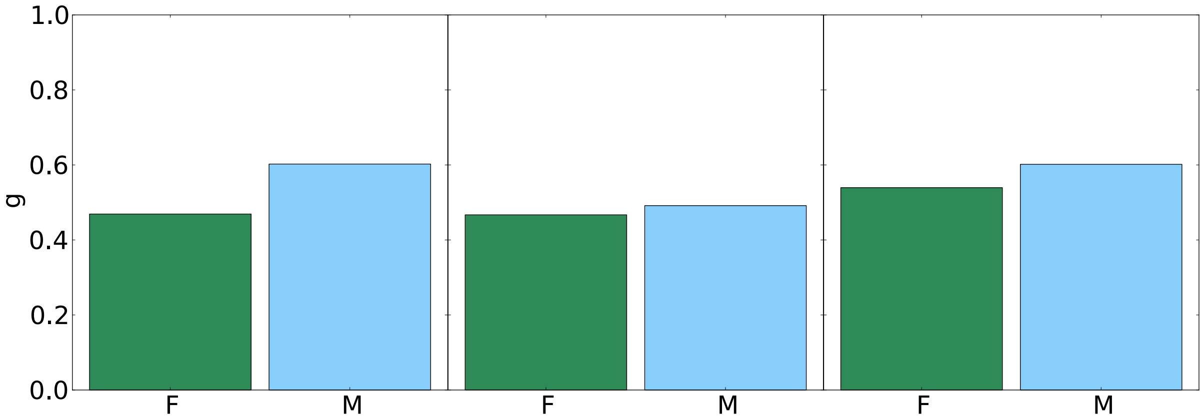

Each panel shows the mean normalized gain, g, for all groups of a given size, N, participating in a given intervention across all three iterations of the experiment. From left to right, the panels show the Ps, Pt and T intervention types. As discussed previously, the tutorial format tends to produce the highest gains and the student-led planetarium shows the lowest. Note, however, that within any given intervention type, there is no statistically significant correlation between group size and mean normalized gain. Smaller groups do not lead to measurably better student test scores for any intervention type.

### *Gender*

We also considered the possibility that conceptual gains might be related to gender. Although we did not expect to find such a relationship, we were surprised to find that it occurred throughout the data. It appeared first in the scores on the pre-test for intervention 1 – the very first test that study participants took and the truest test of their prior knowledge of astronomy. On this test, the mean score among female students was 45%, while among male students the mean score was 55% ($p = 0.0002$).

This difference persisted throughout the experiment, although the differences on subsequent tests were not statistically significant. We found that the average scores of female students were consistently several percentage points lower than those of male students. As shown in Figure 7, we also found that the normalized gains of female students were consistently lower than those of male students.

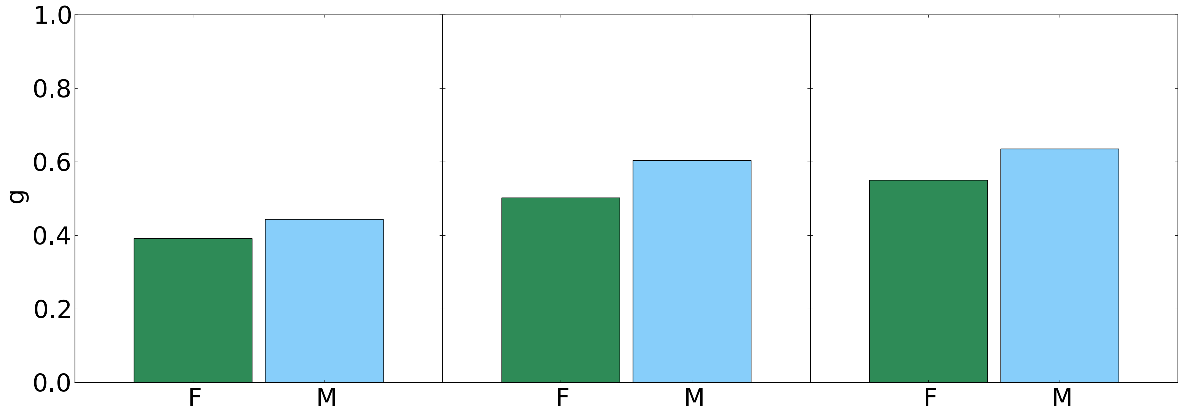
**Figure 7: Normalized Gains as a Function of Student Gender for all three Iterations of the Experiment (1-3, left to right)**




Green bars show the gains, g, for female students and blue bars those for male students. Note that, as discussed earlier, uncertainty intervals are not shown because the data are not Gaussian. The only statistically significant difference is in the first intervention, where p = 0.005.

These differences between male and female students persist across all intervention types. As shown in Figure 8, in all three intervention types, the normalized gains for male students are higher than those for female students.

**Figure 8: Normalized Gains as a Function of Student Gender for all three Intervention Types: Ps (left), Pt (centre) and T (right)**




Green bars show the gains, g, for female students and blue bars those for male students. Note that, as discussed earlier, uncertainty intervals are not shown because the data are not Gaussian.

### *Statistical Significance of the Quantitative Results*

Of primary interest is whether there were statistically significant differences between the intervention types. Table 1 shows the pairings of intervention types for which statistically significant differences in the mean normalized gains were found. In each case, the approach that led to the higher normalized gains is identified.

**Table 1: Statistically Significant Differences among Intervention Types**

| | Iteration 1 | | Iteration 2 | | Iteration 3 | |
|---|---|---|---|---|---|---|
| | **Ps** | **T** | **Ps** | **T** | **Ps** | **T** |
| **Pt** | 0.2 | **0.013 (T)** | **0.01 (Pt)** | 0.84 | 0.79 | 1.0 |
| **Ps** | - | **0.00017 (T)** | - | **0.0017 (T)** | - | 0.70 |

For each iteration of the experiment, we compare the normalized gains from one intervention type to those from each of the other types. In each cell, we indicate p-values for these comparisons, where our threshold of statistical significance is taken to be $p \leq 0.05$. In comparisons with statistically significant p values (bold text), the better of the two interventions is indicated in parentheses.

The results for iteration 1 of the experiment show that tutorials produced higher normalized gains than either student-led or TA-led planetarium shows. The two types of planetarium show were indistinguishable for that iteration of the experiment. These results are likely explained by the fact that the tutorials were twice as long as the planetarium shows during iteration 1, giving students more opportunities to learn prior to taking the post-test.

In iteration 2 of the experiment, tutorials were again found to produce higher normalized gains than student-led planetarium shows, but their advantage over TA-led planetarium shows disappeared. In this iteration, the two types of planetarium show were no longer equivalent: the TA-led ones produced higher gains. By the third iteration of the experiment, all intervention types produced gains that were statistically indistinguishable from one another.

In general, we did not observe statistically significant trends in the normalized gains associated with gender, group size, or the identity of the TA, nor among combinations of these factors. So, for example, we did not observe that smaller group sizes led to meaningfully higher gains for any intervention type, nor for students of either gender.

As might be expected, there were a few cases where a particular TA produced slightly higher gains than other TAs for the same intervention type, but no strong statistically significant trends (for example, that TAs who produced high gains in one intervention type tended to produce low gains in another).

The finding that male and female students had significantly different scores on the pre-test across all intervention types in iteration 1 was among the most statistically significant in our analysis. This was the very first test of the students' knowledge of astronomy, administered at the beginning of the semester, before the students had received much instruction. This difference in the scores decreased over the course of the semester, suggesting that, while female students started off with less knowledge of astronomy, they tended to catch up to their male peers over the course of the semester. This conclusion is borne out by the observation that, by the third iteration of the experiment, the differences between the normalized gains of male and female students, while still present, had fallen below the threshold of statistical significance.

## Qualitative Findings

Our qualitative findings derive from the focus group discussions. As described earlier, we incorporated this qualitative element into our experimental design so that we could measure how students felt about planetarium shows as part of an overall strategy of "highly engaged" teaching. Students were asked questions not only about their experience in the planetarium, but also about other "engaging" aspects of the course. After performing a thematic analysis of our focus group transcripts, we identified the following themes as the most significant:

1. Students' **attitudes** toward the course and the teaching techniques used
2. Students' level of **engagement** in response to different teaching techniques used
3. Students' **learning experiences**, emphasizing the *process* of learning
4. Students' **conceptual gains**, emphasizing the *outcomes* of the learning process
5. **Gender**-related differences in students' approach to and experience in the course
6. Students' impressions of their **TA's** influence on their experience
7. **Suggestions** made by the students to improve the planetarium portion of the experience, or other "engaged" parts of the course

We will treat each of these themes separately from one another and from the quantitative data for now, postponing our attempt at overall synthesis until the next section.

### *Attitudes*

Students' attitudes toward the planetarium were mixed: those who had participated in TA-led planetarium shows were moderately positive about their experience, finding it comparable in usefulness and enjoyment to the tutorial experience and generally superior to the student-led planetarium experience. Some students expressed reservations about the self-directed aspects of the student-led planetarium shows. They were not, in general, confident in their abilities to control the planetarium itself, nor to answer questions using the planetarium in a self-directed way. This led to overall mixed feelings about the planetarium experience: students found it interesting and useful to a degree, but they generally preferred a more traditional style of instruction, with a TA at the front of the room and less onus on the students themselves to direct the experience. It is worth noting that most of these students were first-year university students, the vast majority of whom would never before have participated in a self-directed learning experience of the type we were offering.

Some students questioned the motive for offering activities outside the classroom at all:

> "I also think because it's outside of the classroom setting nobody takes it seriously, so that's why no one cares, like they don't equate that with school work."

"I think it has to do with time, and what is this doing for my career?"

Despite their sometimes dubious attitudes toward time spent in the planetarium, most students found the planetarium itself to be engaging. They appreciated the novelty of the experience and the freedom of the student-led planetarium shows. Many students commented that they found the experience of moving through the universe "amazing" even if they were not sure it helped their learning. Some students were even reasonably sure that the planetarium had not helped them learn more effectively than other methods, but enjoyed the planetarium solely as an engagement tool:

> "Well, I thought it was enjoyable. I didn't learn anything that the textbook or the lectures didn't tell me, but I'm glad they were part of the course."

One might intuitively expect the students to report that the most engaging shows were the student-led shows, in which they got to take the controls and direct their own learning. However, this was not the case. Not only did students have an overall higher impression of the TA-led shows, they found them more engaging too – on par with tutorials and other highly engaged aspects of the course, such as lecture clicker quizzes and interactive online assignments.

### *Engagement*

Students generally found the planetarium experience engaging and, in fact, they identified increasing their engagement as the primary benefit of attending a planetarium show. They tended to value the planetarium for its potential to stimulate wonder, to help them come to terms with the scale of the cosmos, and to engage them in a way not possible with other media:

> "The first planetarium show I went to was amazing for me."

> "I learned that [the universe] was very big and there were many stars."

> "I just think it is a more engaging experience. Being inside the planetarium is more engaging than studying a book."

However, the clear consensus among the students was that the planetarium was more useful for enhancing engagement than for teaching. One reason cited repeatedly for this opinion was that too little time was spent in the planetarium for them to find it pedagogically useful. Students often expressed a desire to spend more time in the planetarium, but on their own terms. For example:

> "The planetarium though, I really like that…I felt was really helpful, but having said that, I did feel it was too rushed."

> "I thought it was really cool. I'd like to have one in my basement. It's really nice to look at."

> "Maybe the planetarium was a little bit [less enjoyable than the tutorials], because it was too short and it was over the whole semester." (Here the student is referring to the planetarium shows being sparsely scattered through the semester, rather than weekly, like the tutorials.)

Students tended to consciously decouple their *attitudes* toward the planetarium learning experience from its potential to *engage* them. They believed that, while it was not always an effective learning tool in itself, it was an engaging experience that drew them more deeply into the course material. This decoupling could be quite pronounced, almost to the point of cognitive dissonance. For example, the same student who made the above comment about the planetarium being "more engaging than studying a book" also said:

> "It wasn't that interesting for me personally. I would rather have been in my tutorial."

This student reconciled her views by noting that, while she found the planetarium itself engaging, she did not like the particular TA who presented her show, nor the instructor-driven format of the show she attended (this particular student did not attend a student-led show and so could not comment on that experience). Other students said that they would have found the planetarium shows more engaging had they been longer and more frequent.

Some students' experiences in the planetarium were conditioned by their expectations, which derived from visits to or knowledge of other larger planetariums. Some typical responses included:

> "I was really excited for it. I thought it was going to be a huge massive thing on the top of the building, but it actually is just this small little blow-up igloo, but it actually is still really cool. I actually enjoyed it."

> "I remember when I was studying in grade 9 or 10…and the school brought us to the planetarium show…It was very impressive at that time because we were so little, and you could just see the whole universe, so I think that was a very powerful personal experience."

These students tended to find our smaller teaching planetarium a little disappointing, yet also reminiscent of their earlier planetarium experiences. To improve the engagement aspect of the students' experience, clearly a larger, more awe-inspiring facility would help.

### *Learning Experience*

Students' impressions of their learning experiences in the planetarium were highly polarized, particularly regarding the student-led shows. Most students felt that the student-led planetarium shows provided poorer learning experiences than either TA-led planetarium shows or tutorials. They often complained about the lack of structure, citing peers who "did things they weren't supposed to do." Other students expressed a corresponding insecurity about participating in the student-led shows:

> "I didn't touch [the controller] because it just looked intimidating. I thought I was going to do it wrong."

> "I was more worried about clicking the right button rather than learning what was being covered in the show."

Another common theme was that students felt "thrown into" the experience of a student-led show. They were very sceptical of taking charge of their own learning. They strongly preferred to be led through the experience by a TA. They were made uncomfortable by the hands-off approach of the TAs in the student-led shows. During that style of show, the TAs were instructed to give prompts to help the students answer their own questions rather than answering them explicitly for them. Clearly, this made some students uncomfortable.

> "It was pretty clear to me that when a TA runs [the show], it is a lot more useful because, when you have questions, they actually will answer them."

> "I think the teacher-led, or instructor was better, because if you do it yourself you don't learn anything, and you could just go to Google instead."

Still, a minority of students felt equally strongly that their learning experience in the planetarium was very positive:

> "I thought the planetarium was extremely helpful… The TA guided it, and then we could ask questions as they went along and we had the clicker questions throughout it, and I thought it was 10X more helpful than all the tutorials combined."

> "I just had one [planetarium show] this year, but, I think we should have more, because it was probably one of the best things about this course, but we just had a lack of time."

These results likely reflect the diversity of the students' learning styles.

Finally, some students felt that the high level of immersion provided by the planetarium helped them focus on the topic at hand:

> "It just seemed like what was covered in the planetarium was more focused on the knowledge in the questions, whereas in the classroom, it is easier to not focus, not pay attention."

### *Conceptual Gains*

Students were asked to speak about the effectiveness of certain course activities in increasing their understanding of astronomy. They had very positive responses to the in-class clicker quizzes, online assignments, lectures, tutorials and online interaction. They were moderately positive in their assessments of the instructor-led planetarium shows, but they had a negative overall view of the effectiveness of the student-led planetarium shows.

Students' positive responses tended to focus on the usefulness of the planetarium for communicating spatial relationships:

> "I liked it. I liked being able to navigate through space and see the distance, and like seeing the relative sizes of things."

> "It didn't really help except that it enforced the fact that these are the planets and this is the way they are aligned."

> "…it did help a little bit, such as how the planets are, how they look in relation to each other…"

> "I think, how I saw that they are disc shaped, how they are aligned, I guess it helped me to remember. In the textbook, it explains this and I think that is good too, but it is helpful to see it." (In response to a question about the ways in which the planetarium was more effective than the lectures)

A significant proportion of students had more negative feelings about their conceptual gains as a result of attendance at planetarium shows, though they often contrasted these with remarks about appreciating the planetarium experience nonetheless:

> "I enjoyed it but I didn't find it really changed anything or that I learned anything from it. I guess I'm sort of neutral. I still liked it but it didn't change anything."

> "I liked the planetarium. I liked it, although I didn't feel like I learned anything, but I liked it as a show. Not that I actually learned any new thing or that we did anything that was fun but I do like the planetarium."

Some students simply had an overall negative impression of the planetarium, expressing a strong preference for tutorials instead:

> "It wasn't that interesting for me personally, I would rather have been in my tutorial."

> "I just want to skip the planetarium and go to my regular tutorial, because it's really pointless, and I just don't see why we would do that."

### *Gender*

There was very little discussion during the focus groups of the relationship between gender and experience in the course. However, male students tended to offer more praise for the planetarium. Most of the most negative comments came from female students. The relatively small number of students involved in the focus groups (a few dozen) makes it hard to ascertain the significance of this observation.

### *TAs' Influence over the Planetarium Experience*

Students were typically quite vocal about the effect of the TA in shaping their experience in the planetarium. Many students praised their TAs for helping them along in the student-led shows:

> "I remember I was in the student-led, and everyone had a chance to take control and [the TA] was asking us to come up with a consensus to answer each question… She was guiding us what to do. I think that was really useful, because she was really encouraging us to talk to each other."

> "The [TA], he really knew his stuff. He was trying to explain everything he could."

Students also drew attention to the critical role of the TA in the other intervention types, as illustrated by this typical remark:

> "I think the tutorials were really helpful, I mean my TA, she was really helpful. If you didn't understand a concept, she would spend the entire tutorial helping you understand it. I found my TA was really, really good with that, so I found the tutorials really helpful."

Still, some students expressed a lot of frustration with their TA, particularly as regarded their teaching abilities or classroom management skills, as illustrated by this comment:

> "Personally my TA I don't think knew what she was doing at all, I mean I think she understood the material, but her teaching ability was really not good, and I don't think she understood how she should have conducted the class at all."

### *Suggestions*

The focus group participants were very forthcoming with their suggestions on how to improve the planetarium experience to make it both more engaging and more pedagogically useful. They were particularly interested in changes to the student-led shows. Broadly speaking, their suggestions focused on improving group dynamics, increasing the frequency of their exposure to the planetarium, spending longer in the planetarium during each visit, and having a more structured or scaffolded experience while there.

Several students commented that they would have appreciated the student-led planetarium sessions more if they had known the people they were working with. They found it difficult to work together with unfamiliar students, particularly when the equipment was also unfamiliar and the time allowed was limited. They would have preferred more time to ‘gel’ as a team before facing uncertain challenges.

The length of the shows was a primary concern, though many students had difficulty deciding if they wanted the shows to be shorter or longer. The consensus seemed to be that, with the shows structured as they were, they would not want them longer (mainly because they didn’t think they were effective and didn’t want to spend more time on them as they were), but with some tweaking, they would want the shows to be longer.

> “Oh the length of the planetarium, in terms of being 20 minutes, was fine, because I don’t think I could handle anymore in that room, and that atmosphere. But if the content was better I would probably say that was too short.”

> “If they do improve it, which is what I think will happen, what they are trying to do, then I think 20 minutes would be way too short because you feel like I want more.”

Some students thought the experience would have been better if it had either been more rigidly scaffolded for them or more individualized and free-form:

> “Yeah it needs to be refined. It just is too controlled. Like I liked it, I wouldn’t mind having that at my disposal at all times.”

> “Well I think that if it was just me sitting in the planetarium, and I had the time to go in and do what I want, it would be cool and I would be able to see just what interests me.”

Students tended to emphasize that, for the planetarium experience to be beneficial, it would need to be more tightly integrated with the course and would have to occur more frequently. They did not like that the content of the planetarium shows was not necessarily related to what they were learning in class in the same week.[5] One student summarized the issue as follows:

> “I feel like they should have, instead of just one or two shows, they should have one every week, so that the content of the show could be what was being covered that week, as opposed to going back a month, and learning the things you learned a month ago.”

Lastly, some students objected to the physical setup of the planetarium. Our particular planetarium uses an inflatable dome; to keep it inflated, a fan must run constantly. This creates a level of background noise that some students found oppressive:

> “I don’t think I could handle any more in that room, and that atmosphere.”

[5] Scheduling constraints inherent to dividing a large class up into small planetarium groups and ensuring that a statistically significant number of students all had comparable experiences made it impossible to keep the planetarium shows in perfect sync with the lectures.

> "I found it a bit difficult to focus because of the sound of the [fan] and sometimes the voice of the TA is low and I can't really hear it that much."

# Conclusions and Recommendations

Based on our results, we can answer our three research questions fairly succinctly: interactive planetarium shows do improve students' understanding of astronomical concepts, but in our particular execution, they produced slightly poorer gains than either planetarium shows led by expert TAs or discussion-based tutorials of equal duration. Despite that, students generally find that planetarium shows increase their engagement with astronomy and they see some promise in the interactive style of planetarium show.

More specifically, we conclude as follows:

1. Planetarium shows are not inherently superior to tutorials for increasing students' short-term conceptual gains. This is clear from Figures 2-4 and the summary presented in Table 1. In fact, there is slight evidence that, when equal time is allowed for both intervention types, tutorials tend to produce higher conceptual gains.

2. First-time non-science astronomy learners experience significant difficulty when asked to use a planetarium in a self-directed way, particularly when their peers are present (and presumed to be exerting peer pressure). This is true even when they are given a simplified control interface customized to help them answer certain pre-set questions. Our TA-led planetarium shows produced marginally higher normalized gains and were favoured by focus group participants.

3. Loosely structured inquiry-style planetarium activities do not work well as an occasional teaching tool. Students feel strongly that, for the activities to be effective, they would need to spend more time in the planetarium, preferably in small groups of familiar peers.

4. Planetarium shows lasting 20 to 30 minutes are not long enough for students to have a satisfying learning experience, unless the show is guided by an experienced presenter. Even then, they would prefer a longer tutorial to a shorter planetarium show.

5. Despite their overall negative assessment of planetarium shows as learning tools, students felt that planetarium shows were important tools for engagement and for illustrating certain spatial relationships not easily presented in either static images or movies.

6. Female students tended to have slightly lower pre-test scores (10% lower at the start of the course) and had marginally lower normalized gains following all three of our intervention types. These results formed a consistent pattern, but did not typically have strong statistical significance when interpreted separately.

7. We see no strong correlation between group size and normalized gains for any of our three intervention types.

From these conclusions, we can suggest several courses of action for other instructors wishing to include planetarium shows as part of their courses.

First, focus on engagement. Unless very great pains are taken to scaffold the planetarium experience, a planetarium show will not necessarily be of any more help to an introductory non-science student than simply talking through concepts (as in a tutorial). However, most students find planetarium shows inspiring and engaging.

Second, when using planetarium shows as teaching tools as part of a large course of non-science majors, it is likely important to schedule many planetarium hours per student. Students take time to acclimate to the planetarium experience; they might initially find it bewildering if not given sufficient time to adapt. In the ideal case, students would have multiple exposures to the planetarium, giving them time to build up competence with the control interface before trying to assimilate astronomical concepts using that interface. Finding the TA resources to run so many shows for such a large number of students will be a significant challenge for most instructors at large universities.

Third, keep in mind that planetarium shows might also have pedagogical uses unknown to the students and not specifically tested. Note that the students in our experiment had strongly negative views of the student-led shows, even though these were not typically statistically significantly worse than the TA-led shows, and often similar to the tutorials. We think that further research into the specific differences between teaching a concept to a small group of students (perhaps even individuals) in a planetarium vs. in a pure discussion-based format would be beneficial. It is entirely possible that planetariums are good for things we have not yet considered or measured.

Finally, the value of a rigorous TA training program cannot be understated. At our institution (which we believe to be typical), TAs receive only a few hours of training, mostly related to policy rather than pedagogy. Any further training provided to teaching assistants must be subtracted from their very limited paid contractual hours. As such, it is extremely difficult to both train the TAs to use the planetarium well and to allocate sufficient hours for them to provide shows to all of the students in a large class. This scheduling tension was perhaps the biggest problem with our approach. It meant that the teaching assistants conducting the planetarium shows had to do so with relatively little training and that there could be relatively few planetarium hours offered overall. Some of our TAs had enough background in inquiry-based techniques to make the most of the short time available for student-led shows, but others struggled. The students noticed: they felt that the TAs were underprepared and the shows were too short.

The best solution to this problem would be to allocate more TA hours to training and yet more to giving planetarium shows more frequently. Practically, this would be problematic. Any increase in training or time spent in the planetarium would have to be deducted from other hours in the teaching assistants' contracts. This makes giving effective, small-group, interactive planetarium shows for large classes very challenging. Another option we are considering is to run several planetariums in parallel, so that students could work alone or in pairs, sharing small planetariums for more time, collectively supervised by a single TA. We hypothesize that having 20 students spread over ten planetariums, supervised by a single TA (and perhaps a system of computerized prompts), would likely lead to greater conceptual gains (though possibly lower engagement) than having 20 students all grouped in one planetarium with a single TA. With the declining costs of planetarium equipment, this sort of approach might soon be feasible. However, before implementing such a strategy, it would be helpful to perform a smaller-scale study on the potential effectiveness of interactive planetarium shows under similar circumstances (small number of students, longer period of access to the facility).

We also recommend further exploration of the issue of how students of different genders learn astronomical concepts. We have detected a small bias in favour of male students, both in pre-test scores and normalized gains. It would be extremely valuable to know if this observation is replicable in other contexts.

# Bibliography


Aikenhead, G., & Michell, H. (2011). *Bridging Cultures: Indigenous and scientific ways of knowing nature.* Toronto: Pearson Canada.

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

Baker, D. (2002). Where is gender and equity in science education? *Journal of Research in Science Teaching*, *39*(8), 656-663.

Bao, L. (2006). Theoretical comparisons of averaged normalized gain calculations. *American Journal of Physics*, *74*(10), 917-922.

Bardar, E. M. (2006). Development and Validation of the Light and Spectroscopy Concept Inventory. *Astronomy Education Review*, *5*(2), 103-113.

Brazell, B. D., & Espinoza, S. (2009). Meta-analysis of Planetarium Efficacy Research. *Astronomy Education Review*, *8*(1), 1-11.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research.* Thousand Oaks, CA: Sage.

Deming, G. (2002). Results from the Astronomy Diagnostic Test National Project. *Astronomy Education Review, 1*(1), 52-57.

Edelson, D. C. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching, 38*(3), 355-385.

Hake, R. (2002). Lessons from the Physics Education Reform Effort. *Conservation Ecology, 5*(2), 28. Retrieved from http://www.consecol.org/vol5/iss2/art28/

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, *53*, 1043-1055.

Hodson, D. (2009). *Teaching and Learning about Science: Language, theories, methods, history, traditions and values*. Rotterdam, The Netherlands: Sense.

Mallon, G. L., & Bruce, M. H. (1982). Student achievement and attitudes in astronomy: An experimental comparison of two planetarium programs. *Journal of Research in Science Teaching*, *19*(1), 53-61.

Plummer, J. (2008). Students' Development of Astronomy Concepts across Time. *Astronomy Education Review*, *1*(7), 139-148.

Prather, E. E., Rudolph, A. L., & Brissenden , G.(2009). Teaching and learning astronomy in the 21st century. *Physics Today, 62*(10), 41-47.

Roth, W. M., & Calabrese Barton, A. (2004). *Rethinking Scientific Literacy.* New York: Routledge Falmer.

Songer, N. B. (2007). Digital resources versus cognitive tools: A discussion of learning with technology. In S. K. Abell & N. G. Lederman (eds.), *Handbook of Research on Science Education* (pp. 471-491). Mahwah, NJ: Lawrence Erlbaum Associates.

Tashakkori, A., & Teddlie, E. (eds.). (1998). *Mixed Methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Taylor, I., Barker, M., & Jones, A. (2003). Promoting mental model building in astronomy education. *International Journal of Science Education, 25*(10), 1205-1225.

Winberg, T. M., & Berg, C. A. (2007). Students' cognitive focus during a chemistry exercise: Effects of a computer simulated pre-lab. *Journal of Research in Science Teaching, 44*, 1108-1133.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario