UNIVERSITY OF CAMBRIDGE
Faculty of Education

ISSN 2043-8338

Journal of Trainee Teacher Education Research

"As clear as night and day". Using models to aid students' conceptual understanding of space and the Solar System in a Year 8 science class

Andrew Derrett

(PGCE Secondary Science (Physics), 2019-2020)

email: andrewderrett@live.co.uk

Abstract

Scientific explanations for basic astronomical phenomena such as the cause of night and day and seasonal variation are notoriously difficult for students to conceptualise, resulting in widespread misconception. Effective teaching of space and the Solar System often suggests the use of models as effective teaching tools. This case study, spanning six astronomy lessons with a Year 8 class, therefore investigates the impact of using seven common teaching models on enabling students to change their conceptions through implementation of a Collated Astronomy Test (CAT). The study also investigates how these students perceived the models in respect of their learning of astronomical phenomena, and provides a basis upon which possible results may guide future teaching of the Earth-Sun-Moon system and the Solar System.

https://doi.org/10.17863/CAM.83617

“As clear as night and day”. Using models to aid students’ conceptual understanding of space and the Solar System in a Year 8 science class

Andrew Derrett

Introduction

Why does the Sun appear during the day and disappear at night? Why is summer warmer than winter? In which direction does gravity act? These questions, and many others, describe everyday phenomena for which most secondary-school students are familiar, but have scientific explanations that are non-intuitive and difficult to conceptualise (Parker & Heywood, 1998). It is little wonder then, that student explanations of the most basic astronomical observations are fraught with misconceptions that may persist unresolved throughout their education and into adult life (Stahly, Krockover, & Shepardson, 1999; Jones, Lynch, & Reesink, 1987; Baxter, 1989; Comins, 1998). Indeed, it is frequently posited that the field of astronomy has just as many, if not more, fundamental misconceptions associated with it than with any other scientific discipline (LoPresto & Murrell, 2011).

Recommendations for teaching practice, in astronomy as well as in wider science education, frequently document the use of *models* as essential pedagogical tools in the promotion of conceptual change (Taylor, Barker, & Jones, 2003; Barnett, Keating, Barab, & Hay, 2000; Chittleborough & Treagust, 2009). Under the assumption that effective science education hinges on helping students develop models of the natural world (Henze, Van Driel, & Verloop, 2008; Etkina & Gentile, 2006), the manner in which students may construct, view and evaluate these models can be argued to be just as significant to their cognitive development as the scientific idea they strive to explicate (Treagust, Chittleborough, & Mamiala, 2004; Chittleborough & Treagust, 2009).

This case study therefore critically investigates the use of models in aiding conceptual understanding of space and the Solar System, with a particular emphasis on how the students *perceive* these models in respect of their evolving knowledge. A number of different models were implemented during a teaching sequence of six lessons delivered to a Year 8 class as part of a PGCE teacher training course whilst on professional placement. This study proposes how data could be collected and analysed in order to determine the effect of these models on both the students' conceptual understanding, and how these students view the models with respect to their learning of space.

This paper begins with a review of relevant literature, substantiating the current understanding of these areas and corroborating these into the research questions to be investigated in the present study. It then proceeds with a justification of the proposed case study methodology and data analysis strategies, before concluding with a discussion of possible results and subsequent implications for teaching practice with regard to the use of models in astronomy classrooms.

Literature review

This section reviews the research that has been conducted around students' a) perceptions of models in the learning of science, and b) conceptual difficulties in astronomy. It will also define the key terms to be used in this paper.

On the definition, classification and taxonomy of models

The term 'model' has a wide variety of meanings in both everyday life and in academic disciplines (Gilbert & Osborne, 1980). Due to diversity of types, use and implications of models in science, a convergent definition for a model is necessarily problematic, reflecting conflicts in epistemological commitments (Shen & Confrey, 2007). A large and evolving set of definitions now exist for a 'model' (Chittleborough, Treagust, Mamiala, & Mocerino, 2005). However, for the purposes of this work, a thorough discussion on the *nature of models as a concept* is of little significance in the determining their practical use in astronomy. Thus, Gilbert and Boulter's (2000) succinct definition is adopted here: a model is a 'representation of an idea, object, event, process or system'. The defining characteristics are that models:

- represent a 'target'
- involve 'relations' between the model and the target; interpretations or simplifications that make them inexact
- have a purpose (e.g. making predictions, testing, teaching).

Constructing a consistent typology for models is equally demanding (Gilbert & Boulter, 2000), and classifications vary depending on the nature of the research. Boulter and Buckley's (2000) typology has been widely adopted, and will therefore be used to categorise the models used in this study.

An important distinction is made here between models as defined above and *mental models* which are "internal, cognitive representations used to reason about phenomena, and to describe, explain and predict" (Boulter & Buckley, 2000, p. 120). For example, a *model* of the Sun – Earth system could be a globe and a lamp, but the *mental model* could be a small, rotating Earth orbiting a central, large, spherical Sun at a great distance.

Views on models

There is a large body of literature supposing that effective teaching of science requires students to develop, analyse and evaluate mental and scientific models (Gilbert & Osborne, 1980; Taylor et al, 2003; Barnett et al, 2000; Torres, Moutinho, Almeida, & Vasconcelos, 2013). Additionally, there are a number of holistic studies that have investigated students' views about the purpose, nature and utility of these models during this process (Chittleborough & Treagust, 2009). Instruments have been created that aim to probe and assess students' perceptions about a range of modelling features, and these will be adapted for the present study (Aikenhead & Ryan, 1992).

In these studies, some students show an impoverished understanding of the model concept, intimating that they must be "exact replicas in every way except size" or be "close to the real thing" (Treagust, Chittleborough, & Mamiala, 2002, p. 359). On the other hand, some studies show that students have a well-developed idea of the purpose of models as explanatory tools, the multiplicity of models, and the changing nature of models (Chittleborough & Treagust, 2009). There are range of other factors that these authors also reveal as significant to students; predominantly the *dynamicity* - whether the model is 2D or 3D - and the *presentation* – what the model allows them to 'see'. Chittleborough et al (2005) consistently showed that, with growing age, students' managed to develop a more detailed appreciation of the role of models in both the scientific process *and* the learning process, a distinction which younger students, and indeed some teachers (Shen & Confrey, 2007), fail to make.

While these studies illuminate students' general perceptions of model concepts, these findings are difficult to apply in specific learning contexts; their limitation is inherent in their generality. It remains unclear to what extent these findings with regard to students' views about models are generic, or how they may be modified in the context of specific topics – astronomy in this instance.

Epistemological comments and the existence of misconceptions

This work does not seek to investigate in depth epistemological considerations around the Nature of Science or implications for students' perceptions of astronomy as a subject, for a plethora of studies exist to that end (Deng, Chen, Tsai, & Chai, 2011). However, any study that examines students' understanding of scientific concepts necessitates an appreciation of the processes by which they learn and the theories that support it.

The predominant theory that describes how children learn has its origin in Piaget's (1936) early work from which the *constructivist* theory of cognitive development was formulated. The fundamental tenet upon which this theory is based is that children assimilate new experiences or information into a pre-existing explanatory framework, whereby they are forced to confront discrepancies between what they know and what they discover, and adjust their ideas accordingly (Driver, 1983).

One of the attractive features of this constructivist theory is that it provides a basis upon which children's frameworks of natural phenomena may result in them developing *misconceptions* (Driver & Easley, 1978). Here, culturally transmitted information (Taylor et al, 2003) is assimilated with existing knowledge to develop a pseudo-logical framework that suffices as an explanatory model but is inconsistent with the scientific idea.

These ideas can be succinctly structured as a progression from intuitive to synthetic to scientific mental models (Agan, 2004). Intuitive models are those based on students' existing explanatory presuppositions (for example, a geocentric model of the Solar System). When the scientific model (heliocentricity) is then assimilated, a synthetic model may be produced that constitutes a misconception (such as Earth's solar orbit as the cause of night and day).

Many terms are used for the word *misconception.* For the purposes of this paper, it will be used synonymously with *alternative framework* and defined with reference to Vosniadou's (1992, p. 536) classification as "explanations of natural phenomena which are frequently different from the currently accepted scientific explanations and which tend to be resistant to change". Here it is noted that this definition precludes misconceptions as simply 'wrong ideas' since, as Barnett et al (2000, p. 135) helpfully state, "understanding of many astronomy concepts are frequently embedded within a larger structure". Therefore, this definition better accommodates the processes by which students develop alternative frameworks in an astronomy context in particular.

Astronomical misconceptions

Instruction of space and the Solar System at secondary-school level usually encompasses a relatively small number of scientific concepts that have remained moderately unchanged over recent decades of teaching (Bailey & Slater, 2003). Lelliott and Rollnick (2009) produced a relevant and substantial review of over 100 published papers in Astronomy Education Research (AER) from 1974 – 2008. They found that five "big ideas" constituted over 80% of the published research, listed in Table 1.

1	Conceptions of the Earth
2	Gravity
3	The day-night cycle
4	The seasons
5	The Earth-Sun-Moon system

Table 1: Lelliott and Rollnick's (2009) five "big ideas" in AER

There is an underlying consensus that teaching of these astronomical concepts is often hindered by the frequency with which students have developed a number of misconceptions around the topics involved (Lelliott & Rollnick, 2009). Since Nussbaum's (1979) seminal work into the misconceptions that children held about the Earth as a cosmic body, a large body of work has surfaced that addresses students', adults' and teachers' conceptions of: the reason for day and night, seasonal variation, the cause of lunar phases and more (Comins, 1998; Comins, 2003; Dunlop, 2000; Bailey & Slater, 2003).

The results of such work culminate in lists of common misconceptions that are held by students of astronomy (LoPresto & Murrell, 2011). Consideration of relevant literature has enabled the researcher to produce a list of nine core misconceptions that are considered in this study, shown in Table 2.

A1	Students struggle to appreciate the relative distances between objects in the Solar System
A2	Students struggle to appreciate the relative sizes of objects in the Solar System
A3	The Lunar phases are caused by the Earth's shadow
A4	Seasons are caused by the distance of the Earth to the Sun
A5	Day and night occur because a) the Earth orbits the Sun or b) the Sun orbits the Earth
A6	There are other stars in our Solar System besides the Sun
A7	'Up' and 'down' directions are absolute. Gravity pulls all objects 'down'
A8	There is no gravity in space
A9	All planetary orbits are circular

Table 2: Nine common astronomical misconceptions

A question remains about how the nature of teaching and learning astronomy lends it to such ubiquitous misconception. An important notion is raised by Barnett et al (2000, p. 135), in that astronomy is difficult to conceptualise correctly "because it requires students to gain an abstract understanding of dynamic relationships and events that take place in 3-D space". A modelling approach then becomes a strategy inviting research, since it can be seen that the core difficulty embedded within astronomy – visualising 3D processes – is very well aligned with Gilbert & Osborne's two "key functions" of models:

> (1) Models enable a simplified version of a phenomenon to be produced and therefore concentrate attention on special features of that phenomenon.
>
> (2) They stimulate investigations, supporting visualization of a phenomenon and imaginative projection.
>
> Gilbert & Osborne, 1980, p.6

Thus, Barnett et al's (2000) comment regarding visualisation and these two functions build a strong case that models are likely to be very effective in astronomical instruction, and that such a strategy warrants research.

Complementary to this, Jones et al (1987, p. 44) note that "children are obliged to conceptualize their environment and … they often do so with a simplistic but self-consistent logic". Baxter (1989) strengthens this stance with the realisation that children autonomously construct their own mental models long before they receive any formal education. Thus, children enter the classroom with a variety of internally formulated frameworks explaining what they observe, but that the logical steps they have taken in constructing these are incompatible with the accepted scientific ideas (Driver, 1983).

Nussbaum's (1979) work provided one of the first in-depth looks at children's misconceptions of an astronomical phenomenon – specifically the nature of gravity on Earth. This study built on Piaget's (1929) founding articles on how children conceive astronomical phenomena, which have been influential for decades. Through a set of semi-structured interviews, Nussbaum (1979) established that children generally hold one of five 'notions' of Earth. These progress from a primitive 'flat-earth' notion through some alternative frameworks to the accepted scientific model.

To emphasise this point, Piaget (1929) intimates that astronomy is conceptually demanding because children are required to overcome their 'egocentric' frame of reference. For example, in order to conceptualise the Earth as a sphere in the cosmos, one has no option but to visualise what their surroundings look like from an observer positioned in outer space: such demands are difficult and non-intuitive for children and adults (Parker & Heywood, 1998).

Nussbaum's (1979) work acts as a starting point in the elicitation of children's misconceptions, but it is not clear what the precise criteria for these notions are, or whether these frameworks had arisen through the students' miscomprehension of the interview questions (Panagiotaki, Nobes, & Potton, 2009). Therefore, Vosniadou and Brewer (1992) aimed to establish whether these notions were held in a consistent manner through a similar but more rigorous interview process. They interviewed 60 students of 6 to 11 years, and used a precisely defined scoring guide to elicit and categorise consistently-described notions. In this way, a very similar set of notions for the Earth were established, with conceptual change through childhood also documented.

This study – with its robust methodology – has built up a strong case for children's Earthly misconceptions, and numerous other studies (Diakidoy, Vosniadou, & Hawks, 1997; Vosniadou & Brewer, 1994) have appeared to confirm them. On the other hand, Schoultz, Säljö and Wyndham (2001) note that students' notions change drastically when a globe is introduced in the interview as a visual prop. That said, using a globe as a prop for the interview may have provided the child with visual clues that could, as Nussbaum (1979) says, "interfere" with their natural thinking, and thus care must be taken when using their study to analyse students' intuitive mental models.

The nature of the Earth constitutes a large fraction of AER. However, much has been achieved with regard to mental models of other astronomical concepts such as the cause of day and night, the seasons, and lunar phases. In each case, students' have been shown to hold a pattern of synthetic models that are inconsistent with the scientific model. Most common of these are; Earth-Sun distance as the cause of seasonal variation, a geocentric model as the cause of day/night, and the Earth's shadow as the cause of the lunar phases (Dunlop, 2000; Barnett & Morran, 2002). It is noted here, as a matter of interest, that a child's development from a primitive geocentric flat-earth model to a scientific heliocentric model parallels the historical development of the Solar System, which is deemed to be a useful teaching framework (Jones et al, 1987; Baxter, 1989).

Assessment of conceptual understanding in astronomy

There has been a growth in the use of 'concept inventories' or diagnostic tests since the creation of the Force Concept Inventory by Hestenes and Halloun (1985). They are research-based tests, usually multiple-choice, that probe student understanding of a physics topic. They have particular utility in measuring the effectiveness of teaching and are carefully developed using an iterative process of gathering ideas from students and experts, and revising the test (Sands, Parker, Hedgeland, Jordan, & Galloway, 2018).

There have been a wide range of diagnostic tools developed in astronomy education, each with slightly differing foci (Slater, 2015). Some of these have been developed to assess students' holistic understanding of astronomy concepts (Hufnagel, 2002) while others target more specific areas (Williamson, Willoughby, & Prather, 2013). The majority of these tests have been developed for introductory college courses in the United States. Therefore, their suitability *individually and in isolation* is extremely limited for the present case; that being a Year 8 class in the English education system. Much of the content in these tools is targeted for a level of understanding far more comprehensive than this group are expected to have developed.

Therefore, this section summarises the tools used for the current study. The manner in which they are *collated*, *modified*, and *simplified* for the purposes of ensuring the tests are suitable for the students in question is justified in the methods section. They are all designed to be 'pre-post' surveys; administered once before an intervention, and once after, in order to measure conceptual change. Figures 1 to 5 illustrate example questions from each of the tools discussed below.

ADT 2.0

The Astronomy Diagnostic Test (ADT) was developed by Hufnagel and released up to a version 2.0 in 1999 (Hufnagel, 2002). It constitutes a 33-question pencil and paper multiple choice survey. A strength of this test for the present study is that it a) was developed "to include only concepts recognizable to most high-school graduates", and b) "avoid(s) jargon" (Hufnagel, 2002, p. 1). Therefore, many of the questions in this test are directly relevant for KS3 students.

Moreover, the possible answers provided in this test were designed using "only phrases that students volunteered in their interviews" (Hufnagel, 2002, p.2). Hence the language used in this test is the natural language for students and does not include technical vocabulary. Furthermore, the test was designed to provide a holistic assessment of the students' astronomical understanding, and so many of the questions naturally elicit the misconceptions provided in Table 2. This enables straightforward mapping between question and misconception, allowing for richer analysis of students' results with respect to conceptual change or otherwise.

11. Compared to the distance to the Moon, how far away is the Space Shuttle (when in space) from the Earth?
    A. Very close to the Earth
    B. About half way to the Moon
    C. Very close to the Moon
    D. About twice as far as the Moon

Figure 1: ADT 2.0 sample question (adapted from Hufnagel, 2002)

AMS

The Astronomical Misconceptions Survey (AMS) was developed by LoPresto and Murrell (2011), and has the advantage of being a more recent construction than the ADT. Its development centred around the elicitation of "major" astronomical misconceptions, and therefore is appropriate as a diagnostic tool in the present study.

22. What causes the phases of the Moon?
    a. The Moon's shadow on Earth.
    b. Earth's shadow on the Moon.
    c. The Sun and Earth's shadows on the moon.
    d. None of the above-no shadows are involved.

Figure 2: AMS sample question (adapted from LoPresto & Murrell, 2011)

MOSART 5-8 Astronomy Test

The National Science Foundation of the United States funded a Harvard University-based project entitled *Misconception-Oriented Standards-Based Assessment Resources for Teachers* (MOSART, Sadler et al, 2010). As part of this project, they developed diagnostic tests for a wide use in the sciences, including astronomy. There is a limited literature-base supporting their creation or analysing their effectiveness, but the 'Astronomy 5-8' test consists of material directly relevant to the posed misconceptions, since these were designed for a target group of similar age to the present case. Although these tests were constructed with a focus on the astronomy standards established by the United States' high-school curriculum, the content is directly comparable to that of the National Curriculum (Department for Education, 2014).

13. Our solar system contains:
    a. One average star
    b. Several stars spread across space
    c. One older, dimmer star, and one younger, brighter star
    d. Three stars
    e. No stars

Figure 3: MOSART 5-8 sample question (adapted from Sadler et al, 2010)

TOAST

The Test for Astronomy Standards (TOAST) serves as the most recent astronomy assessment tool, designed by Slater (2015). It inherited questions from previous assessments, which were then modified with the addition of four further questions in line with advice from 28 "experts". This tool has been shown to be reliable, having been used with over 2000 students and published in three peer-reviewed journals.

7. Imagine that the Earth's orbit were changed to be a perfect circle about the Sun so that the distance to the Sun never changed. How would this affect the seasons?
   A. We would no longer experience a difference between the seasons.
   B. We would still experience seasons, but the difference would be much LESS noticeable.
   C. We would still experience seasons, but the difference would be much MORE noticeable.
   D. We would continue to experience seasons in the same way we do now.

Figure 4: TOAST sample question (adapted from Slater, 2015)

NGCI

The Newtonian Gravity Concept Inventory (NGCI) was developed by Williamson et al (2013), and assesses concepts involved in classical gravitation. Other force or gravity concept inventories exist and have shown to be reliable (Sands et al, 2018), but this tool has the advantage of being explicitly aimed at astronomy students and thus tests gravity on an astronomical scale. Hence it is more relevant for the present study. Importantly, this test addresses some concepts that some of the others used exclude, for example in assessing the variation of the gravitational field with mass and distance. Therefore, incorporating it for the present study ensures the evidence collected in addressing all relevant misconceptions is more comprehensive.

6. An astronaut floating in her Earth-orbiting spacecraft …
    a. experiences no gravitational force from Earth.
    b. still experiences a gravitational force from Earth.
    c. experiences a force from the spaceship that counters the gravitational force from Earth.

Figure 5: NGCI sample question (adapted from Williamson et al, 2013)

Models in astronomy

There are few pieces of research that indicate the impact of using a modelling approach in astronomy on student understanding. Henze et al (2008) and Torres et al (2013) describe the success of such an approach in improving *teacher* subject knowledge, and others describe general modelling approaches (Barnett & Morran, 2002; Taylor et al, 2003). There remains little analysis that identifies the effect of using astronomical models, or how students perceive them during their use. The researcher notes that this may be because using models as part of space teaching has become increasingly engrained in light of a move towards a modelling approach in science education (Halloun, 2006).

Research questions

The following two research questions (RQs) result from consideration of the above literature, and aim to probe further the effectiveness of using models in aiding students' conceptual understanding of space, and also what the students' perceptions of these models are.

RQ1: Do instructional approaches using models improve students' conceptual understanding of space and the Solar System?

RQ2: How do students perceive models with respect to their learning of space and the Solar System?

Therefore, the results from this investigation may be used to evidence recommendations for practice in terms of modelling strategies, and also to elucidate the processes underpinning students' understanding of these models and the role they play in instigating conceptual change.

Teaching rationale

Context

This study took place in a secondary school of approximately 1000 students in the east of England. The school educates a mixed student cohort from ages 11 to 16, derived from a small but diverse urban catchment area. The school is non-selective, and in its most recent Office for Standards in Education inspection it was rated ‘Outstanding’ in overall effectiveness and all other categories.

The class studied in this work was a small Year 8 class, consisting of 13 students. The group had a range of previous attainments, but all students had target GCSE grades between five and seven, establishing this as a lower-attaining group.

The researcher taught six lessons - split over four 100-minute periods - as part of the school’s Key Stage 3 (KS3) ‘Space’ scheme of learning. Due to timetabling, the class was taught two further lessons during this time by one of the science teachers who shared teaching the class. These additional lessons were ‘Refraction’ and ‘Telescopes’; these were deliberately chosen as being of limited relevance to the posed research questions and thus interfered little with the modelling research.

Justification of research methodology

This works aims to answer two research questions with regard to the teaching and learning of astronomical phenomena. There are a number of approaches by which the two posed research questions could be investigated, but the methodology adopted here is that of *case study.*

There are a number of advantages that this broad approach has in the current study. RQ1 is inherently outcome-driven – the outcome being the promotion of conceptual change in students’ understanding (or otherwise) – which could at first invite an action research study (Denscombe, 2007). However, RQ2 constitutes an investigation into the *underlying processes, links and relationships* by which the students learn. When examining such themes, the wider approach offered by a case study can prove much more illuminating (Yin, 1984).

In this specific case, students’ own perceptions on the role models have played in their cognitive development of the delivered topic are necessarily complex (Smith & Tanner, 2010), and thus an outcome-driven methodology would be less suitable in establishing the subtler mechanisms by which students apply the models in their learning. RQ2 requires consideration of the ‘how and why’ as well as the ‘what’ required by RQ1. As Denscombe (2007, p. 36) puts it, “case studies tend to be holistic… (emphasising) relationships/processes rather than outcomes…”.

What’s more, having identified a limitation in this study – that being the involvement of another teacher in the learning sequence – the case study approach better relieves this *by inviting a variety of types of data and data collection* (Yin, 1984; Denscombe, 2007). As such, incorporating multiple sources of data means a case study approach in this instance can lead to more reliable conclusions for the posed research questions.

To strengthen this stance, case studies are best chosen when investigating situations in their ‘natural setting’ (Denscombe, 2007; Bell, 2005). The specific case for this work – that being the Year 8 class – had already been taught by the researcher for a number of lessons in their previous science topic, and so were already familiar with the teacher. And, more importantly for this study, *traditional instruction of astronomy very often involves or recommends the use of models for teaching purposes.* In their wide-scale AER review, Lelliott and Rollnick (2009, p.1791) note that “there is considerable support in the literature for modelling activities”. Seeing as this principle is heavily reflected in previous studies, it is strongly suggested that using models as a pedagogical tool in space lessons represents a natural setting for these students, and thus the case study approach is justified.

The models

This study involved seven models that were used as part of the lesson sequence, shown in Table 3 (next pages). These models were chosen (a) in order to address all misconceptions presented in Table 2 earlier and (b) for pragmatic reasons, indicating models which are constructed and implemented with relative ease, using materials found in most secondary schools. In line with the defining features of all models (Lee, Chang, & Wu, 2015), each model has its associated *targets* and *relations* identified, and are characterised using Boulter and Buckley’s (2000) typology as set out in Table 4 (following Table 3 below) by means of the following scheme:

1. The first letter denotes the *mode* of representation: concrete (C), verbal (Ve), visual (Vi), mathematical (M) or gestural (G). For simplicity, ‘mixed’ modes are excluded.
2. The second letter denotes the *dynamicity* of the model: static (S) or dynamic (D).
3. The number denotes the *dimensionality,* either two or three dimensional (2 or 3).

For example, the gravity well is a *concrete*, *dynamic*, *3-D* model and is thus assigned type CD3.

Model		Description	Type	Target(s)	Relation(s)
M1	Human orrery	The students are tasked with organising themselves into a scale model of the Solar System. The students themselves represent the Sun, planets and other objects in the Solar System, such as current positions of manmade spacecraft. On the chosen scale, Mercury is positioned 10 cm from the Sun, Neptune 8.08 m, and the Voyager I Spacecraft 38.0 m.	GD2	- The structure and order of celestial objects in the Solar System. - The relative distance between objects in the Solar System.	*Students* as *planets*. *Walked paths* as *orbits.* *Floor* as *3D space.*
M2	Playground Solar System	The students are tasked to establish, using researched data, their own appropriate scale with which to represent the Solar System. Then, using chalk, they sketch out their scale diagrams on the playground surface.	ViS2	- The structure and order of celestial objects in the Solar System. - The relative distance between objects in the Solar System.	*Drawn pictures* as *planets.*
M3	Plasticine planets	The students receive a moderately-sized lump of plasticine. Then, using Noel-Storr's (2012) steps, they are tasked with dividing up the plasticine in the appropriate manner. This enables them to place replicas of the planets next to each other, which are accurate with regard to their relative size. These are placed in a line on dark A3 paper.	CS3	- The order of celestial objects in the Solar System. - The relative size of objects in the Solar System.	*Plasticine* as *planets within the Solar System.* *A3 Card* as *3D space.*
M4	Gravity well	This is coordinated as a demonstration. 100 g weights represent a heavy object such as the Sun, and are placed in the centre of a stretched lycra sheet, representing the fabric of spacetime. This enables gravity as a bending of spacetime, its variation with mass and distance, and the shape of planetary orbits to be modelled and visualised using marbles. A clear description is found at Science Learning Hub (2019).	CD3	- Gravity as spacetime-curvature. - Variations of the force-due-to-gravity with mass and distance. - Shape of planetary orbits.	*Lycra sheet* as *fabric of spacetime.* *Weight(s)* as *Sun.* *Marbles* as *planets.* *Dynamic motions* as *orbits.*
M5	Globe/lamp model	This is coordinated as a demonstration with a globe and a 100W bulb, representing the Sun. Fictitious people are placed on each hemisphere of the globe, and a discussion is held about what season it would be for each of them. The globe is taken on a revolution around the Sun, enabling seasonal variation throughout a year to be modelled.	CD3	- The cause of the seasons on a planetary scale.	*Lamp* as *Sun* (incl. *lamplight* as *sunlight*). *Globe* as *Earth.*

Model		Description	Type	Target(s)	Relation(s)
M6	The tilted lamp	This is coordinated in pairs. Each pair is given a lamp, and a strip thermometer. They are tasked with establishing how the temperature, as indicated by the thermometer, depends on the tilt of the lamp. Then they are encouraged to use this model to explain seasonal temperature variations, using the ideas of heat energy and surface area.	CS3	- The cause of seasonal temperature changes.	*Card* as *surface of the Earth.* *Lamp* as *Sun* (incl. *lamplight* as *sunlight*).
M7	Polystyrene balls/lamp model	Each student receives a polystyrene ball, representing the Moon, and positions themselves around a central 100W bulb, representing the Sun. The lights are turned off. By a fixed rotation, the appearance of Lunar phases due to the amount of a half-lit Moon visible from Earth is modelled. The National Science Teaching Foundation (2014) provides a clear description of this model as used in the classroom.	CD3	- The cause of the Lunar phases.	*Head* as *Earth.* *Eyes* as *Earthly observer* *Lamp* as *Sun* (incl. *lamplight* as *sunlight*). *Static rotation* as *Lunar orbit.* *Lamp revolution* as *Earth orbit.*

Table 3: The seven models deployed in the study, with associated type, targets and relations

Mode of representation	**Description of the mode of representation**
Concrete	Material models; e.g. a scale model of the Solar System
Verbal	Written or spoken models, descriptions or explanations
Visual	Models that are seen, such as drawings or videos
Mathematical	Models that use formulae or equations to represent a phenomenon
Gestural	Models that are movements of the body: e.g. using hands to describe how the planets orbit each other

Table 4: Boulter and Buckley's (2000) modes of representation, constituting part of their model typology

Methods, methodology and implementation

Ethics

This research was carried out in accordance with the British Educational Research Association (BERA) guidelines for research (BERA, 2018). All participants were informed of the nature of the research project and what form the lessons would take. This decision was taken upon ethical grounds, although Bell (2005) and Denscombe (2007) note that this provides a weakness for the case study approach: those being researched may behave differently knowing that they are being witnessed as part of a study. Although this does imply that the class may have been placed outside of their 'natural setting', it was decided for this group of students that this effect was of minimal significance.

The class teacher, joint science teacher and Head of Science - acting as 'gatekeepers' (BERA, 2018) - were fully informed before the study took place. It was decided that no part of the study was detrimental to the students' progress. Students were reassured that the study did not constitute formal assessment.

In keeping with school policy, letters were sent to guardians allowing them to consent to the recording of their students in interviews. Consent was received for 11 out of the 13 students. Students were assured that all data would be anonymised in the final paper, and the Faculty of Education ethics form was signed and approved by the Subject Lecturer.

Data collection

The study's research questions aim to determine whether modelling pedagogies are effective in the teaching of astronomy, and what the students' perceptions on the role these models are. For this reason, data collection strategies were chosen to determine students' conceptual understanding before, during and after the lesson sequence, in addition to their overarching views and experiences throughout.

A limitation of this type of study is that the lack of control group means it is difficult to ascertain the *cause* of any measured change in students' understanding of taught concepts. This is reflected in social science research as the problem of identifying *causal validity*. As Schutt (1995, p. 51) puts it: "solutions are neither easy nor perfect: We always have to consider critically the validity of causal statements that we hear or read". Thus, it was decided in this study that simply measuring students' conceptual understanding via diagnostic tests was insufficient to reliably address the two research questions or to formulate *valid* conclusions as to the effectiveness of modelling pedagogies.

Therefore, in this study, exit passes (quantitative and qualitative), individual interviews and a focus group were implemented in addition to a diagnostic test. Taken together, these multiple data sources are deployed in order to substantiate evidence that could reliably address the research questions, although causal validity remains a limitation.

Table 5 illustrates the data collection strategies used in addressing each of the two Research Questions, each of the strategies being described and justified below.

Research Question	Exit Pass	Diagnostic Test	Focus Group	Semi-structured Interview
RQ1: Do instructional approaches using models improve students' conceptual understanding of space and the Solar System?	✓	✓		
RQ2: How do students perceive models with respect to their learning of space and the Solar System?	✓		✓	✓

Table 5: The research questions and associated data sources

Creating a 'Collated' Astronomy Test (CAT)

Best practice in administering diagnostic tests would dictate that a single one of the tools described in the literature review would be completed by each student as a pen-and-paper exercise both before and after the lesson intervention (Madsen, McKagan, & Sayre, 2017). However, given the noted problems of target audience and literacy, such a strategy is not suitable for this lower-attaining group (Henderson & Wellington, 1998).

As such, the five tools described were combined into a single *Collated Astronomy Test* consisting of 20 multiple choice questions, the construction of which followed the following principles:

- Each question should be mapped onto one of the misconceptions determined in Table 2. This ensures measurement of conceptual change can be achieved and any patterns identified.
- Each question should not be modified, unless it is deemed that the original wording or content is unsuitable for the target group.

- All misconceptions should be mapped onto at least one question.
- The test should be as succinct as possible, with no repeated questions, to ensure the scope is restricted to "crucial issues related to the research…avoiding any superfluous detail" (Denscombe, 2007, p. 161).
- Students should be given one minute per question, in line with recommended time limits from the tests constituting the CAT.

Following these principles ensures the resulting CAT is a robust tool in assessing the target group's conceptual understanding of *relevant* astronomical phenomena. However, it should be noted that – since each test is validated as a single instrument – this approach reduces overall reliability. That said, since each *individual* question is constructed based on student-responses and common misconceptions, the 'distractors' (wrong answers) are carefully selected, and so each question acts as an independent, reliable, assessment of a specific concept (Sands et al, 2018; Hestenes & Halloun, 1985). Hence collating these questions as described, it is argued, does not significantly reduce overall reliability.

The CAT, however, is not exempt from the wider limitations of these assessments. Smith and Tanner (2010) note that conceptual understanding can be obscured by jargon, and a reliance on closed-ended, multiple-choice questions necessitates that they primarily assess content knowledge over conceptual understanding. Despite these limitations, concept inventories are seen as effective tools in assessing the impact of instructional approaches (Sadler et al, 2010), and so the CAT was applied in conjunction with the following alternative data sources.

Exit Passes

Written student responses to questions for students to complete at the end of a lesson – frequently labelled e*xit passes* – have become a popular tool for teachers for both formative assessment and in the promotion of student and teacher reflection (Leigh, 2012). In this study, they were administered at the end of each lesson, for two purposes:

Purpose 1: To assess their understanding of the lesson's astronomical phenomenon.

Purpose 2: To probe perceptions of the model with respect to their learning of astronomy.

Adopting these two purposes for exit passes naturally exploits the two areas where exit passes have been shown to be most effective (Jeyaraj, 2019). Purpose 1 constitutes formative assessment of conceptual understanding, and as such provides data that addresses RQ1. Purpose 2 constitutes qualitative student reflection of their views of the model, and as such provides evidence for RQ2.

In this study, exit passes with two questions were constructed for each lesson to satisfy both of these purposes. In each instance, the first question was designed to test conceptual understanding. The second question, which remained unchanged throughout the study, required some open reflection: "What did you think of today's model? Did it help you understand more about space?" (see Figure 6 later). Authors have recommended the use of such open-ended questions in exit passes to facilitate reflection and guide future teaching (Leigh, 2012; Jeyaraj, 2019).

To improve the reliability of data with regard to purpose 1 of the exit passes, particular attention was given to the first 'concept' question to ensure it provided reliable assessment of their understanding. As such, these questions were adapted from diagnostic resources provided by the University of York's Best Evidence Science Teaching (BEST) project, which is a research-informed curriculum development scheme. Produced resources draw on a wide base of science education literature, and the diagnostic questions in 'Big Topic: Earth in Space' were deemed suitable for use as exit passes (BEST, 2019). An example exit pass, issued after lesson four, is shown in Figure 6.

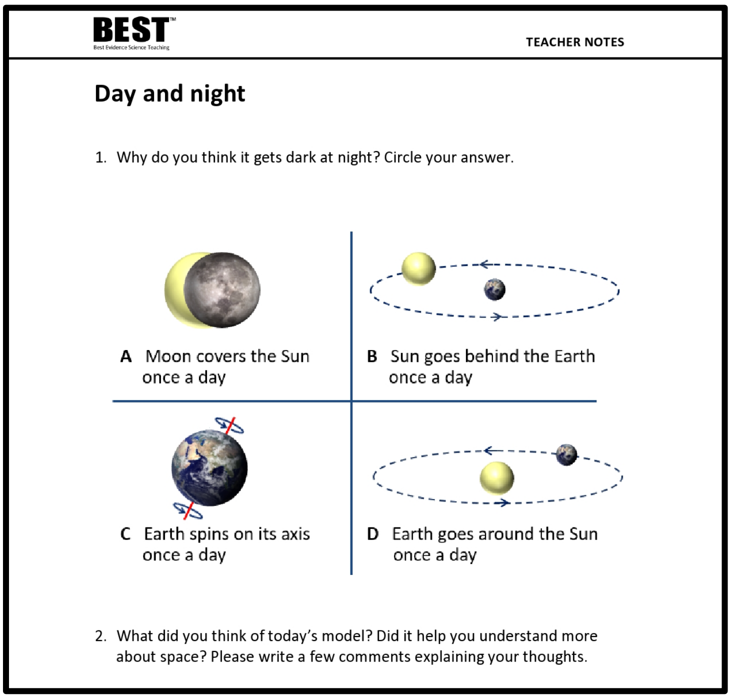

Figure 6: Exit pass for Lesson Four, an adapted diagnostic question from BEST 2019 materials

Focus Group

To supplement the data received from the CAT tests and exit passes, a 30-minute focus group with the class was coordinated after the lesson sequence. Here, acting as moderator, models were introduced as ‘stimuli’ for the discussion (Denscombe, 2007). During discussions, data was recorded in the form of observational notes, with particular attention to comments regarding students’ perceptions of the models. The purpose of this focus group was to substantiate evidence for RQ2 by gathering opinions about the teaching models, and what they ‘meant’ for the students.

The focus group provided an open environment in which students could offer their views in a non-confrontational atmosphere. Discussions between students were allowed to evolve naturally without a strict agenda and without dictating the sequence of talk (Denscombe, 2007). However, in order for discourse to consistently yield relevant evidence for RQ2, the researcher used ‘hinge questions’ to help guide the discussions. The questions adopted for this purpose were drafted from the tool *My Views of Models in Science* (VOMMS) created by Chittleborough and Treagust (2009) in a study that aimed to develop a holistic understanding of the students’ views on the nature of ‘teaching’ and ‘scientific’ models. The tool is therefore applicable to the subset of models that form the focus of this study. These adapted ‘hinge’ questions, borne out of Chittleborough and Treagust’s (2009) study, are shown in Figure 7.

1. Are these models just representations of how things work, or accurate duplications of reality?
2. Are these models the only way to explain the science, or are there other ones too?
3. Do you think scientists like to focus on just one model when investigating space, or do they prefer to use lots at the same time?
4. Will these models change in future years or stay the same?
5. What do you think led scientists to accept and use these models for space?

Figure 7: Hinge questions used in focus group and interviews, adapted from Chittleborough and Treagust (2009)

The questions were adapted for several reasons. Firstly, to ensure they were appropriate for the age-group. Secondly, to frame them in an ‘open’ manner, whereby the students are presented with a range of contrasting views but are also able to voice their own (Denscombe, 2007). And also, as will be elucidated in the analysis section, in order to frame the questions in such a way so as to promote discussions that *reveal the students’ perceptions* about the models. This enables the researcher to identify and record, via observational notes, emergent themes with relative ease (for example, comments with regard to *dynamicity, multiplicity, explanation, purpose* and more).

The strength of conducting such a focus group for the present study is that it allows for students to hear, share and discuss a range of contrasting or concurrent ideas (Bell, 2005). In this way, students are able to hear others' points of view as well as articulating their own, which has the benefit that it "reveals the reasoning and underlying logic used by participants" (Denscombe, 2007, p. 179). In this fashion, rich data can be collected that addresses RQ2 insofar as it is likely to reveal *why* the students perceive the models as they do, as well as what those perceptions are.

Furthermore, this lower-attaining group was highlighted as containing students with notable literacy difficulties in reading and writing. As such, conducting a focus group has the added benefit of allowing those students to voice their responses verbally, thereby bypassing any literary barriers (Cyparsade et al, 2013). Thus, data collected in this fashion is likely to *supplement* as well as *support* existing data from the CAT and exit passes. However, it should be noted that this focus group was to be conducted with 13 students, which is above the recommended limit for such a strategy, and leads to confident students dominating the discussion and the possibility of not all participants contributing (Denscombe, 2007; Bell, 2005). To address this limitation, semi structured interviews were also used.

Interviews

Short audio-recorded semi-structured interviews were conducted at the conclusion of the study with the 11 students who consented. As with the focus group, these were coordinated with attention to the perception that these students had developed with regard to the presented models, and how they may have enabled them to learn. For simplicity and consistency, the same hinge questions were used as in Figure 7, but with scope embedded within the interview for the student to expand, explain and build on their response. Despite Schoultz et al's (2001) suggestion of using the models as prompts, it was decided on grounds of practicality not to bring each physical model into the interview. Advantages of these interviews are that a) they ensure responses from all participants are received and b) audio-transcriptions allow for more detailed data analyses, as described in the next section (Denscombe, 2007). The purpose of the interviews was to reveal any relations between an individual student's conceptual understanding, and their perception of the relevant models, which may guide recommendations for practice.

Collated Astronomy Test (CAT) - pre

Lesson	Misconception(s) targeted	Model(s)	Exit Passes
1. Scale model of the Solar System	A1, A6	(M1) Human orrery (M2) Playground Solar System	✓
2. Plasticine planets	A2	(M3) Plasticine planets	✓
3. Gravity and Orbits	A8, A9	(M4) Gravity well	✓
4. Day and night	A5	(M5) Globe/lamp	✓
5. The seasons	A4	(M5) Globe/lamp (M6) The tilted lamp	✓
6. The Lunar Phases	A3	(M7) Polystyrene balls/lamp	✓

Collated Astronomy Test (CAT) - post

Focus Group

Semi-structured Interviews

Figure 8: Temporal structure of the case study, including data sources

Figure 8 summarises the structure of the case study, highlighting data sources. The key strength of this approach is that it allows for *methods triangulation.* That is, “checking out the consistency of findings generated by different data collection methods” (Patton, 1999, p. 1193). Such a strategy not only seeks to find sources that yield the ‘same’ result in each instance, but also discrepancies, which may also prove illuminating (Patton, 1999). For example, in this study, exit pass responses and CAT results may indicate a consensus that students have changed their frameworks of a particular misconception by providing complementary results, or indeed that the extent of conceptual change might be more limited than each source may have suggested in isolation. Similarly, observations from the focus group can be used in parallel with interview transcripts and exit passes to reveal, support and counter what students’ perceptions on the posed models may have been.

For example, if a specific student answered a question mapped to a particular misconception correctly in both the relevant question in the post CAT *and also* in the respective lesson’s exit pass, then it can be concluded more confidently that a student had succeeded in formulating their understanding using the accepted scientific model. The same principle applies in the reverse situation whereby a student answers both incorrectly and thus still harbours some alternative frameworks. This strategy addresses the limitation that all diagnostic tests have – including the CAT – in that students may answer correctly via a ‘guessing’ method that does not necessarily indicate correct conceptual understanding. As such, this approach enhances reliability and validity (Scaife, 2004).

Data analysis and discussion

This section explores some potential findings from the data sources with respect to each of the research questions, and describes how a coding procedure could be adopted for the qualitative data sources. These possible findings are linked to some recommendations for practice and possible future work in the final section. Inferential statistics are not used in this study due to the small sample size provided by this class (Opie, 2004).

RQ1: Do instructional approaches using models improve students’ conceptual understanding of space and the Solar System?

This research question would be addressed by descriptive quantitative analysis of the CAT tests in conjunction with question one of the exit passes. Data would be analysed in order to present results from the whole class as well as individual students, and presented in line with the literature supporting the development of diagnostic tests. (Madsen et al, 2017).

Raw Gain (RG)	$post - pre$
Normalised Gain (NG)	$\frac{post - pre}{100\% - pre}$
Effect Size (ES)	$\frac{post - pre}{stdev}$

Table 6: Three measures of conceptual change using CAT results (Madsen et al, 2017)

The three measures shown in Table 6 would be calculated. The *raw gain* (RG) is the crudest indication of what the students had learnt during the study. However, because this was the first teaching this class had been delivered on space, it is likely that the pre-test scores would have been low. Hence, the *normalised gain* (NG) is a better measure of their gain in knowledge compared to their knowledge deficit in the pre-test. However, neither of these measures take into account the spread of students’ scores, and normalised gain has been shown to underreport effectiveness of teaching if pre-test scores are low (Coe, 2002). For this reason, the *effect size* (ES) is a useful measure for students’ learning, by comparing the raw gain to the standard deviation of the students’ scores. An effect size below 0.2

is considered small, 0.5 is a medium effect size, and above 0.8 indicates a large difference (Coe, 2002; Madsen et al, 2017).

To strengthen this data, exit pass (EP) scores, taking into account question 1 only, would be calculated and provided as a percentage, supporting the CAT results. Since exit passes were issued immediately after each lesson, but the CAT post-test after the lesson sequence, this strategy enables consistencies, inconsistences or temporal variations in students' results to be identified.

Such a strategy would permit the researcher to discuss *holistic* findings from the study with regard to the overall effectiveness of the modelling approaches adopted in the lessons. Some possible findings, implications and subsequent recommendations are collated in Table 7. Existing literature on using models in astronomy teaching (Barnett et al, 2000; Henze et al, 2008, Shen & Confrey, 2007) suggests that this study is likely to be an effective strategy in promoting conceptual change. If this is not shown, then there would be some important implications regarding the use of these specific models for low-attaining Year 8 students, and perhaps in astronomy education more widely, which warrant further attention.

Whilst such analysis would illuminate some general implications in light of whole-class results, it is likely that a more detailed examination of specific *misconceptions, clusters of questions,* and *students* will better reveal patterns and guide more explicit recommendations (Madsen et al, 2017). This strategy enables a differentiated analysis of which misconceptions and models may have been successful, and which not. A possible structure for results, implications and recommendations is shown in Table 8, where it is seen that a more nuanced analysis of each of the individual lessons could provide actionable recommendations for the educator of astronomy.

Another strength of this approach is that it may enable a clearer basis upon which future research should be focused. For instance, in the example results, the *educator* may be interested in the use of models M1 and M2 (scale models) in their classroom, but the *researcher* may be more interested in the applicability of model M4 (gravity well) in other contexts where it may be more successful. The existing literature suggests that models M1, M2 and M5 are likely to be effective (Asher, Bailey, Christou, & Popescu, 2006; Schoultz et al, 2001), but the researcher offers no intuition or expectation with regard to the others.

NG	ES	EP	Implications and Recommendations
0.8	0.8	90%	Students showed a very large, consistent increase in conceptual understanding of all taught astronomical phenomena, with very few misconceptions, if none at all. It is concluded that the modelling pedagogies as delivered were very effective at improving students' conceptual understanding and promoting the scientific ideas. Therefore, their implementation as in the current study is strongly recommended.
0.6	0.5	85%	Students showed a notable increase in conceptual understanding in most of the taught astronomical phenomena. However, some students still held some misconceptions after the study. There is some inconsistency between the CAT results and the exit passes, which may indicate that students were reverting back to their alternative frameworks. The modelling approach is shown to be effective overall, but it is not clear where or how it could be improved, and so further work is needed.
0.4	0.3	40%	Students showed a small increase in conceptual understanding of the taught astronomical phenomena. However, there were still widespread misconceptions held by the students, as indicated by both the CAT results and the exit passes. The modelling approaches may have been effective teaching strategies, but there is evidence that students remained confused and that some of the models were of little use or implemented inappropriately.
0.1	0.1	25%	Students showed little or no increase in conceptual understanding of the taught astronomical phenomena. Students were still extensively adopting alternative frameworks. The CAT results indicate very little successful conceptual change, and the exit passes scores are indicative of random guessing. The modelling approaches used were ineffective teaching strategies, and their use as implemented in this study is not recommended.

Table 7: Illustrative class-wide findings with associated implications and recommendations

Misconception	Question(s)	Model(s)	NG	ES	EP	Implications and Recommendations
A1	1, 3, 7	M1, M2	0.9	0.8	95%	Students were highly successful in improving their conceptual understanding of relative distance in the Solar System. The human orrery and playground Solar System enabled students to adopt the scientific scale model of the Solar System, and thus their use in astronomy lessons is highly recommended.
A2	2	M3	0.6	0.5	71%	Most students were successful in understanding the relative size of objects in the Solar System. The plasticine planets model enabled some students to adopt the scientific mental model of the size of planets, but some students' alternative frameworks were resistant to change. There is some inconsistency in the CAT results and exit passes. This model is recommended, but should be adapted to suit all students.
…	…	…	…	…	…	…
A5	5, 8, 17, 19	M5	0.5	0.4	48%	Some students were successful in understanding the scientific cause of day and night. However, many students were still harbouring alternative synthetic models, as indicated by the CAT results and exit passes, which are consistent. This model could be recommended, but more work should be done to establish how its effectiveness can be improved.
…	…	…	…	…	…	…
A9	4	M4	0.1	0.2	27%	Students were unsuccessful in understanding the shape of planetary orbits in the Solar System, with most still believing them to be circular. The gravity well, as implemented, was unsuccessful in promoting conceptual change; the CAT results and exit passes are consistent and indicate little more than random guessing. The model, as implemented, is not recommended.

Table 8: An illustrative structure to analyse conceptual change for specific misconceptions, with associated implications and recommendations

RQ2: How do students perceive models with respect to their learning of space and the Solar System?

This question demands analysis of students' qualitative responses to question two of the exit passes, the focus group and the individual interviews. As with other studies (Chittleborough et al, 2005; Chittleborough & Treagust, 2009; Lee et al, 2015), these responses would be coded into a set of emergent themes, from which student's common, core perceptions could be identified. For clarity, responses would be organised and coded for each data source separately before triangulation.

An existing set of codes

In their creation and analysis of the Students' Understanding of Models in Science (SUMS) instrument, Treagust et al (2002) identified five key themes that consistently emerged in students' interviews, shown in Figure 9.

(1)	Scientific models as multiple representations
(2)	Models as exact replicas
(3)	Models as explanatory tools
(4)	How scientific models are used
(5)	The changing nature of scientific models

Figure 9: Treagust et al's (2002) five key themes, into which qualitative data could be coded

RQ2 constitutes a research question in tight alignment with their study, with the proviso that perceptions in this instance are constrained to *astronomical* models. Hence, these five themes provide a relevant framework into which exit pass responses, focus group observations and interview transcripts could be coded.

Example thematic responses and their implications

(1) Multiple representations

Here, students perceive that different models can be used to represent the same target, but to provide a 'variety of perspectives and appearances'. This notion is most likely to appear in this study in relation to the lessons in which two models were used: the scale model of the Solar System and the seasons.

For example, a response of "I understand why it gets warm in summer now because the globe showed that summer and winter are in different places in the year, and the tilted card showed why that means it's warmer" would indicate a perception that different models can be used to represent the same concept. Indeed, this comment would support the existing literature (Lelliott & Rollnick, 2009) in recommending that students be afforded the opportunity to critically evaluate different models with regard to their strengths and limitations.

(2) Exact replicas

Treagust et al (2009, p. 359) showed that students often have the notion that models must be "close to the real thing" or "exact in every way except size". In an astronomical context, such impoverished understandings are significant hindrances to developing correct mental models, since the very concepts of size and distance are central to an accurate view of the Solar System.

Thus, a response such as "I could see the planets' sizes with the plasticine easily, and they were the same size as the Moon when we did the phases bit" would indicate that the student has perceived both models as being exact replicas, and has not distinguished accurately between the physical models and the realities they represent.

The prevalence of such comments is likely to guide recommendations. If many students were to intimate such perceptions, then a recommendation would be to ensure that when implementing the appropriate models, that students are guided to distinguish the model from its target. Such recommendations are likely to surface because the nature of the Solar System dictates that models *necessarily* cannot be exact scale replicas in terms of distance *and* size simultaneously.

(3) Explanatory tools

Responses such as "the model showed different parts of the Solar System" or "it allowed me to see it more easily" indicate that students perceive that astronomical models hold significant explanatory power. The frequency of these comments and, more importantly, with which model they reference, will heavily guide recommendations for their potential implementation.

For example, if students frequently note that the polystyrene model allowed them to "see how the reflection off the Moon makes you see the phases", then using such a model when teaching Lunar phases would be recommended. This stance would be strengthened, of course, if students were also able to answer the equivalent diagnostic questions accurately, *showing that the model's explanatory power had enabled them to learn effectively*.

(4) How scientific models are used and *(5) Their changing nature*

Previous literature indicates that many students do not see that models have a purpose other than for explanation (Lee et al, 2015). On the other hand, Treagust et al (2002, p. 359) did find that the majority of students agreed that models "can change according to advances in scientific thinking". The qualitative results from this study would be able to determine whether these perceptions persist in an astronomical context. The researcher offers no intuition as to whether or not these findings will correlate to existing research, other than the fact that the structure of the Solar System – having been studied for centuries – may appear resistant to change to students, and thus they could be unlikely to hold the perception that the corresponding models may change over time.

Building robust recommendations

The quantitative and qualitative results from the study could be combined to substantiate teaching recommendations for the use of models in astronomy. For example, students' CAT and exit pass results may suggest a secure understanding of the cause of the seasons. On top of this, exit passes, focus group observations and interviews could indicate that the students deem *multiplicity* and *explanation* as core aspects of their perception of these models (e.g. "I understood it because I could see the more tilted one being less hot", or "the globe showed the tilt and the lamp was the heating bit."). Such results, when taken together, would lead to a recommendation to use *both* the tilted card and globe/lamp model when teaching about seasonal variation *because* it allows students to visualise how both Earth's position in its orbit and its tilted axis generate seasons, *and* it encourages them to develop an enhanced appreciation about the process of modelling in science.

Individual student tracking would be useful in the elicitation of anomalies or patterns that clustered or whole-class results might obscure. As such, a strategy whereby the results from a single student's CAT tests, exit passes, interviews and focus group responses are collected and analysed would enable the educator to establish more precisely the learning process of that student, and thus build up robust, actionable conclusions to guide future teaching. Table 9 provides a possible structure. For example, although student B was generally unable to progress to scientific mental models, they *were* formulating a correct understanding of relative distance. This analysis would allow models to be compared and contrasted. Some authors have criticised multiple-choice tests, suggesting that they underestimate students' knowledge (Dunlop, 2000). Hence, building recommendations using the above procedure can be argued to increase reliability.

	Data		**Student A**		**Student B**	
RQ1	CAT	Pre	40%		30%	
		Post	70%		40%	
		RG	30%		10%	
		NG	0.5		0.1	
		ES	0.5		0.2	
	Mis-conceptions	A1	CAT: 80%	EP: ✓	CAT: 100%	EP: ✓
		A2	CAT: 75%	EP: ✓	CAT: 40%	EP: X
		…	…	…	…	…
		A9	CAT: 40%	EP: X	CAT: 0%	EP: X
RQ2	EP, interview & focus group responses		*"When we were running around we could see how far everyone else was to us … but then the plasticine game showed how big each of us should have been to be right for the planets."*		*"The gravity well wasn't right because there's nothing to support the Sun in space so there's no gravity in space."* *"The lamp thing showed the Earth gets hot because the lamp was close to the strip thing so it's to do with how close it is."*	

Implications	Implications
This student achieved considerable improvement in their conceptual understanding of astronomical phenomena. The student is shown to be formulating their understanding of relative size and distance using the scientific mental model as indicated by the consistent CAT and exit pass results, hence the human orrery and plasticine planets were successful in aiding their understanding. They conceptualise the shape of planetary orbits incorrectly, and so the gravity well was ineffective. The interview responses show that this student perceives the *multiple representations* of models as important as enablers of the *explanatory power* of the scale models and plasticine planets.	This student did not achieve significant improvements in conceptual understanding of astronomical phenomena. They performed poorly in the CAT test and exit passes, indicating little more than guessing throughout. The models were ineffective for this student. There is a notable exception in that this student performed excellently with regard to questions concerning relative distance in the Solar System across both the CAT test and the exit pass. Hence we can conclude that the human orrery and playground Solar System were effective in promoting conceptual change for this student. Their interview responses show that they deem models to be *exact replicas* and this has been a barrier to developing their understanding, as they continue to hold alternative frameworks.

Recommendations	Recommendations
The models should be implemented as in this study, but with a particular emphasis on *comparing* and *contrasting.* The evidence from this student is that they were able to change their initial misconceptions by relating the human orrery and the plasticine planets together and combining them into the scientific mental model. This enabled them to have an accurate understanding of the Solar System after instruction. This approach is recommended in lessons, and such students should be encouraged to voice their views and enable other students to adopt similar frameworks to improve their conceptual understanding.	When implementing the models, emphasis should be placed on models on representations and not as exact replicas. Students should be clear on the limitations of each of the models, making it explicitly what the 'targets' and 'relations' are: this is important in the gravity well. The evidence from this student is that they were not able to connect the physical models to the realities they represent, so a focus on abstracting the models into mental frameworks is recommended.

Table 9: A possible structure by which individual student data could be presented, leading to classroom implications. 'Tracked' data enables patterns and anomalies to be identified

Conclusions, limitations and final remarks

The aim of this study was to investigate the use of models in the learning of astronomy, with a focus on how Year 8 students perceive models in their learning process. Seven models, spanning six lessons, were chosen to address nine core misconceptions students have been shown to harbour about the Solar System. The study proposes how a collated astronomy test, exit passes, interviews and a focus group could be utilised to facilitate data triangulation and produce reliable results with regard to their conceptual understanding and views about models. Possible frameworks for analysing quantitative and qualitative results are provided, building on existing work, and example results are given which would enable the researcher to make some robust recommendations as to the implementation of these models in classrooms. However, the results of this study are limited in their degree of generalisability. Firstly, they do not address to what extent students' prior achievement, instruction, or cultural background may affect their perceptions. There is also the problem of the sample size being small, and the issue of attaching their conceptual change to the modelling pedagogies. This work also does not address how different implementations of each model might yield different levels of conceptual understanding: a more detailed study of individual models would be needed for this.

While this study is likely to support the use of models in astronomy classes, the extent to which recommendations could be applied to other areas of science education may be very limited. Inherent in this is the ubiquity and diversity of modelling approaches in the instruction of science. Indeed, many are proposing an entire pedagogical shift in science from constructivist learning to a modelling theory of learning (Hestenes, 1987; Halloun, 2006). Such theories remain out of the scope of this work, which is limited in its ability to guide far-reaching recommendations as to the use of models in *all* contexts, or to provide a detailed understanding of how models are perceived or applied by students in wider science education. Nevertheless, the conclusions from a study such as this allow the astronomy educator to confidently deploy, assess and evaluate a range of modelling techniques in their classrooms, enabling students to enhance their own understanding of the Earth and its place in the Solar System.

References

Agan, L. (2004). Stellar ideas: Exploring students' understanding of stars. *Astronomy Education Review*, *3*(1), 77-97. https://doi.org/10.3847/aer2004008

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: 'Views on science—Technology—Society' (VOSTS). *Science Education*, *76*(5), 477-491. https://doi.org/10.1002/sce.3730760503

Asher, D. J., Bailey, M. E., Christou, A. A., & Popescu, M. D. (2006). The human orrery: A new educational tool for astronomy. *Astronomy Education Review*, *5*(2), 159-176. https://doi.org/10.3847/aer2006024

Bailey, J. M., & Slater, T. F. (2003). A review of astronomy education research. *Astronomy Education Review*, *2*(2), 20-45. https://doi.org/10.3847/aer2003015

Barnett, M., Keating, T., Barab, S. A., & Hay, K. E. (2000). Conceptual change through building three-dimensional virtual models. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth international conference of the learning science* (pp. 134-141). Mahwah, NJ: Erlbaum.

Barnett, M., & Morran, J. (2002). Addressing children's alternative frameworks of the moon's phases and eclipses. *International Journal of Science Education*, *24*(8), 859-879. https://doi.org/10.1080/09500690110095276

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, *11*(5), 502-513. https://doi.org/10.1080/0950069890110503

Bell, J. (2005). *Doing your research project: a guide for first-time researchers in education, health and social science*. Buckingham: Open University Press.

Best Evidence Science Teaching (BEST). (2019). *Big idea: Earth in space*. STEM Learning - Resources, CPD, STEM Ambassadors and enrichment. https://www.stem.org.uk/best/physics/big-idea-earth-space

Boulter, C. J., & Buckley, B. C. (2000). Constructing a Typology of Models for Science Education. In J. Gilbert & C. Boulter (Eds.), *Developing models in science education* (pp. 41-57). Netherlands: Kluwer Academic Publisher.

British Educational Research Association (BERA). (2018). *Ethical guidelines for educational research*. Retrieved April 13, 2020, from https://www.bera.ac.uk/wp-content/uploads/2018/06/BERA-Ethical-Guidelines-for-Educational-Research_4thEdn_2018.pdf

Chittleborough, G. D., Treagust, D. F., Mamiala, T. L., & Mocerino, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science & Technological Education*, *23*(2), 195-212. https://doi.org/10.1080/02635140500266484

Chittleborough, G. D., & Treagust, D. F. (2009). Why models are advantageous to learning science. *Educación Química*, *20*(1), 12-17. https://doi.org/10.1016/s0187-893x(18)30003-x

Coe, R. (2002). *It's the effect size, stupid: what effect size is and why it is important* [Paper presentation]. Paper presented at Annual Conference of the British Educational Research Association, Univeristy of Exeter, England.

Comins, N. (1998). Identifying and addressing astronomy misconceptions in the classroom. *New Trends in Astronomy Teaching*, 118-123. https://doi.org/10.1017/cbo9780511628993.028

Comins, N. F. (2003). *Heavenly errors: Misconceptions about the real nature of the universe*. Columbia University Press.

Cyparsade, M., Auckloo, P., Belath, I., Dookhee, H., & Hurreeram, N. (2013). Beating the Language Barrier in Science Education: In-Service Educators' Coping with Slow Learners in Mauritius. *Science Education International*, *24*(4), 402-415.

Deng, F., Chen, D., Tsai, C., & Chai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, *95*(6), 961-999. https://doi.org/10.1002/sce.20460

Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd ed.). Buckingham: Open University Press.

Department for Education. (2014). *The national curriculum in England*. Retrieved April 13, 2020, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/381344/Master_final_national_curriculum_28_Nov.pdf

Diakidoy, I., Vosniadou, S., & Hawks, J. D. (1997). Conceptual change in astronomy: Models of the earth and of the day/night cycle in American-Indian children. *European Journal of Psychology of Education*, *12*(2), 159-184. https://doi.org/10.1007/bf03173083

Driver, R. (1983). *The pupil as scientist?* Milton Keynes: Open University press.

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, *5*(1), 61-84. https://doi.org/10.1080/03057267808559857

Dunlop, J. (2000). How children observe the universe. *Publications of the Astronomical Society of Australia*, *17*(2), 194-206. https://doi.org/10.1071/as00194

Etkina, E., Warren, A., & Gentile, M. (2006). The role of models in physics instruction. *The Physics Teacher*, *44*(1), 34-39. https://doi.org/10.1119/1.2150757

Gilbert, J., & Boulter, C. (2000). *Developing models in science education*. Netherlands: Kluwer Academic Publisher.

Gilbert, J. K., & Osborne, R. J. (1980). The use of models in science and science teaching. *European Journal of Science Education*, *2*(1), 3-13. https://doi.org/10.1080/0140528800020103

Halloun, I. A. (2006). *Modeling theory in science education*. Dordrecht: Kluwer Academic Publishers.

Henderson, J., & Wellington, J. (1998). Lowering the language barrier in learning and teaching science. *The School science review*, *79*(288), 35-46.

Henze, I., Van Driel, J. H., & Verloop, N. (2008). Development of experienced science teachers' pedagogical content knowledge of models of the Solar System and the universe. *International Journal of Science Education*, *30*(10), 1321-1342. https://doi.org/10.1080/09500690802187017

Hestenes, D., & Halloun, I. A. (1985). Common sense concepts about motion. *American Journal of Physics*, *53*(11), 1056-1065. https://doi.org/10.1119/1.14031

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, *55*(5), 440-454. https://doi.org/10.1119/1.15129

Hufnagel, B. (2002). Development of the astronomy diagnostic test. *Astronomy Education Review*, *1*(1), 47-51. https://doi.org/10.3847/aer2001004

Jeyaraj, J. S. (2019). Promotion of learner autonomy through entrance & Exit slips. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3486347

Jones, B. L., Lynch, P. P., & Reesink, C. (1987). Children's conceptions of the earth, sun and moon. *International Journal of Science Education*, *9*(1), 43-53. https://doi.org/10.1080/0950069870090106

Lee, S. W., Chang, H., & Wu, H. (2015). Students' views of scientific models and modeling: Do representational characteristics of models and students' educational levels matter? *Research in Science Education*, *47*(2), 305-328. https://doi.org/10.1007/s11165-015-9502-x

Leigh, S. R. (2012). The Classroom is Alive with the Sound of Thinking: The Power of the Exit Slip. *International Journal of Teaching and Learning in Higher Education*, *24*(2), 189-196.

Lelliott, A., & Rollnick, M. (2009). Big ideas: A review of astronomy education research 1974–2008. *International Journal of Science Education*, *32*(13), 1771-1799. https://doi.org/10.1080/09500690903214546

LoPresto, M. C., & Murrell, S. R. (2011). An astronomical misconceptions survey. *Journal of College Science Teaching*, *40*(5), 14-22.

Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Best practices for administering concept inventories. *The Physics Teacher*, *55*(9), 530-536. https://doi.org/10.1119/1.5011826

National Science Teaching Association. (2014). *Moon Phases Demonstration* [Video]. YouTube. https://www.youtube.com/watch?v=wz01pTvuMa0

Noel-Storr, J. (2012). *A scale model of the Solar System using Play-Doh!* STEREO Learning Center. Retrieved April 15, 2020, from https://stereo.gsfc.nasa.gov/classroom/scales.shtml

Nussbaum, J. (1979). Children's conceptions of the earth as a cosmic body: A cross age study. *Science Education*, *63*(1), 83-93. https://doi.org/10.1002/sce.3730630113

Opie, C. (2004). *Doing educational research*. London: SAGE.

Panagiotaki, G., Nobes, G., & Potton, A. (2009). Mental models and other misconceptions in children’s understanding of the earth. *Journal of Experimental Child Psychology*, *104*(1), 52-67. https://doi.org/10.1016/j.jecp.2008.10.003

Parker, J., & Heywood, D. (1998). The earth and beyond: Developing primary teachers' understanding of basic astronomical events. *International Journal of Science Education*, *20*(5), 503-520. https://doi.org/10.1080/0950069980200501

Patton, M. Q. (1999). Enhancing the quality and credibilty of qualitative analysis. *Health Services Research*, *35*(2), 1189-208.

Piaget, J. (1929). *The child's conception of the world*. London: Kegan Paul, Trench & Trubner.

Piaget, J. (1936). *The origins of intelligence in children*. London: Routledge & Kegan Paul.

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: Development and validation of assessment instruments aligned with the K–12 national science standards. *Astronomy Education Review*, *8*(1). https://doi.org/10.3847/aer2009024

Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies*, *3*(1), 173-182. https://doi.org/10.1080/23752696.2018.1433546

Scaife, J. (2004). Chapter 4: Reliability, Validity and Credibility. In C. Opie (Ed.), *Doing educational research: A guide to first-time researchers* (pp. 58-72). London: SAGE Publications Ltd.

Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Heavenly talk: Discourse, artifacts, and children’s understanding of elementary astronomy. *Human Development*, *44*(2-3), 103-118. https://doi.org/10.1159/000057050

Schutt, R. K. (1995). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: SAGE.

Science Learning Hub – Pokapū Akoranga Pūtaiao. (2019). *Exploring physics concepts with a gravity well*. Retrieved April 15, 2020, from https://www.sciencelearn.org.nz/image_maps/76-exploring-physics-concepts-with-a-gravity-well

Shen, J., & Confrey, J. (2007). From conceptual change to transformative modeling: A case study of an elementary teacher in learning astronomy. *Science Education*, *91*(6), 948-966. https://doi.org/10.1002/sce.20224

Slater, S. J. (2015). The development and validation of the test of astronomy standards (TOAST). *Journal of Astronomy & Earth Sciences Education (JAESE)*, *1*(1), 1. https://doi.org/10.19030/jaese.v1i1.9102

Smith, J. I., & Tanner, K. (2010). The problem of revealing how students think: Concept inventories and beyond. *CBE—Life Sciences Education*, *9*(1), 1-5. https://doi.org/10.1187/cbe.09-12-0094

Stahly, L. L., Krockover, G. H., & Shepardson, D. P. (1999). Third grade students' ideas about the lunar phases. *Journal of Research in Science Teaching*, *36*(2), 159-177. https://doi.org/10.1002/(sici)1098-2736(199902)36:23.0.co;2-y

Taylor, I., Barker, M., & Jones, A. (2003). Promoting mental model building in astronomy education. *International Journal of Science Education*, *25*(10), 1205-1225. https://doi.org/10.1080/0950069022000017270a

Torres, J., Moutinho, S., Almeida, A., & Vasconcelos, C. (2013). Portuguese science teachers' views about nature of science and scientific models. *Ninth International Congress on Research in Science Teaching*, 3541-3546.

Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*(4), 357-368. https://doi.org/10.1080/09500690110066485

Treagust, D. F., Chittleborough, G. D., & Mamiala, T. L. (2004). Students' understanding of the descriptive and predictive nature of teaching models in organic chemistry. *Research in Science Education*, *34*(1), 1-20. https://doi.org/10.1023/b:rise.0000020885.41497.ed

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535-585. https://doi.org/10.1016/0010-0285(92)90018-w

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/Night cycle. *Cognitive Science*, *18*(1), 123-183. https://doi.org/10.1207/s15516709cog1801_4

Williamson, K. E., Willoughby, S., & Prather, E. E. (2013). Development of the Newtonian gravity concept inventory. *Astronomy Education Review*, *12*(1). https://doi.org/10.3847/aer2012045

Yin, R. K. (1984). *Case study research: Design and methods*. Newbury Park, CA: SAGE.