Conceptualising the Cosmos: Development and Validation of the Cosmology Concept Inventory for High School

Saeed Salimpour^{1,2,3,4} · Russell Tytler¹ · Brian Doig¹ ·
Michael T. Fitzgerald^{5,6} · Urban Eriksson⁷

Received: 31 May 2021 / Accepted: 21 January 2022 / Published online: 5 February 2022
© The Author(s) 2022

Abstract

Cosmology concepts encompass complex spatial and temporal relations that are counterintuitive. Cosmology findings, because of their intrinsic interest, are often reported in the public domain with enthusiasm, and students come to cosmology with a range of conceptions some aligned and some at variance with the current science. This makes cosmology concepts challenging to teach, and also challenging to evaluate students' conceptual understanding. This study builds on previous research of the authors investigating the methodological challenges for characterising students' cosmology conceptions and the reasoning underlying these. Insights from student responses in two iterations of an open-ended instrument were used to develop a concept inventory that combined cosmological conceptions with reasoning levels based on the SOLO taxonomy. This paper reports on the development and validation of the Cosmology Concept Inventory (CosmoCI) for high school. CosmoCI is a 28-item multiple-choice instrument that was implemented with grade 10 and 11 school students ($n = 234$) in Australia and Sweden. Using Rasch analysis in the form of a partial credit model (PCM), the paper describes a validated progression in student reasoning in cosmology across four conceptual dimensions, supporting the utility of CosmoCI as an assessment tool which can also instigate rich discussions in the science classroom.

Keywords Cosmology education · Concept inventory · Student conceptions · Partial credit model

✉ Saeed Salimpour
ssalimpour@deakin.edu.au

Extended author information available on the last page of the article

Introduction

Cosmology as a field of inquiry has roots in mythology, philosophy, and religion. Cosmology pursues answers to some of the biggest and most fundamental questions about the Universe. As a precision observational science cosmology aims to tell the evolutionary history of the universe and predict its future. Cosmology topics are prevalent and consistent across most curricula at upper secondary level (Salimpour et al., 2020a). The cosmology curriculum, because of the fundamental questions raised about our place in a mysterious universe, has enormous potential to engage the curiosity of students and create rich discussions in the classroom. Simultaneously, these topics encompass complex space–time relations which are innately counterintuitive and require reasoning beyond everyday experiences (Salimpour et al., 2021b, in review).

Conceptions and Reasoning

The history of research into student alternative/misconceptions in astronomy is rich and diverse. Over the years, various misconceptions have been identified and various evidence-based interventions proposed to scaffold students towards conceptual change. The iSTAR database (Slater et al., 2016) contains 186 articles focussing on some aspect of misconceptions in astronomy. Some examples of alternative conception studies include topics such as night/day (e.g.: Vosniadou & Brewer, 1994), seasons (e.g.: Slater et al., 2018), moon phases (e.g.: Trundle et al., 2007), cosmology (e.g.: Prather et al., 2002), size/distances (e.g.: Miller & Brewer, 2010), and astrobiology (e.g.: Offerdahl et al., 2002). One of the underlying challenges in addressing student conceptions is that they are based on intuition grounded in everyday experiences and language (e.g.: Nussbaum & Novak, 1976; Vosniadou & Brewer, 1992, 1994), making them resistant to change (Driver & Easley, 1978). Furthermore, although studies have been extremely valuable in identifying alternative conceptions, there is much work that can be done in characterising the reasoning that underpins these conceptions.

Reasoning is one of the foundations of science, and one of the thrusts of science education has been to “instil the disciplinary habits of the mind of the scientist” (Kind & Osborne, 2017, p. 9). However, as argued by Kind and Osborne (2017), despite the importance and richness of reasoning in science, science education has yet to conceptualise and characterise scientific reasoning in the classroom in a way that is reflective of the epistemic practices of science. Following Driver and Easley (1978) in order to characterise alternative conceptions in a way that allows them to be addressed effectively requires an understanding of the underlying reasoning patterns. Alternative conceptions research has in essence been of two varieties —*nomothetic* and *ideographic* (Driver & Easley, 1978). While nomothetic studies have their place and compare student understanding to a standard, ideographic studies can productively explore the reasoning underpinning students’ conceptions (e.g.: Vosniadou & Brewer, 1994).

Previous Work on Concept Inventories

Concept Inventories (CIs) have gained popularity as a basis for formative and summative assessment processes, particularly in astronomy education. CIs are diagnostic tests consisting of multiple-choice questions designed to explore a student's understandings of a particular construct or a series of constructs/concepts (Bailey, 2009; Sadler et al., 2009; Wilson, 2005). There has been much research into the benefits of using CIs (Bailey, 2009; Wallace & Bailey, 2010), and the methodologies used to create them (Lindell et al., 2007). As of this reading, a range of validated CIs exist in astronomy (Table 1), which have mostly been of the nomothetic type aimed at undergraduate level and focus on levels in conceptual understanding and do not necessarily privilege reasoning.

Validated CIs focussing on cosmology education include those of Wallace (2011), which is aimed at undergraduate students in an introductory astronomy course. The CI consists of conceptual questions on very focussed topics in cosmology: expansion and evolution of the universe, the Big Bang, and the evidence for dark matter in spiral galaxies, which are aligned to the level the concepts are taught at undergraduate level. The questions do explore student reasoning informed by construct (concept) maps for each of the topics; however, that is not the primary focus. The CI is designed to be a pre-/post-test to measure student learning gains via Lecture-Tutorials in undergraduate introductory astronomy. The work of Aretz et al. (2016) focuses on exploring student pre-instructional ideas about the Big Bang Theory, while Aretz et al. (2017) used student responses and the work of Wallace (2011) to refine a construct map about the expansion of the universe. As such, there is no CI in cosmology developed specifically for high school students that privileges a progression in reasoning combined with conceptual and declarative knowledge. Given the complex and counterintuitive nature of much of cosmology, understanding and supporting students' reasoning are important as it helps unpack the complex space–time relations, the epistemic practices of the discipline, and the rich narrative of how we have uncovered the mysteries of the Cosmos. Characterising hierarchies of reasoning can also provide teachers with guidance on how to frame their teaching to effectively address alternative conceptions, in a way that scaffolds students to unpack the concepts in cosmology.

In our previous work (Salimpour et al., 2021b, in review) using student responses from an open-ended survey, we identified alternative conceptions which aligned with previous studies (Aretz et al., 2016; Hansson & Redfors, 2006; Prather et al., 2002; Trouille et al., 2013; Wallace, 2011), but extended these to identify the underlying reasoning patterns. Looking deeper at the alternative conceptions revealed that they were linked to three fundamental reasoning challenges associated with the following:

1. Navigating spatial and temporal relations over enormous scales
2. The counterintuitive nature of cosmological concepts
3. Intuition based on everyday language and experience

Table 1 A collection of concept inventories in astronomy, together with their purpose and target level

Concept inventory	Purpose	Target level	Reference
Test of Astronomy Standards (TOAST)	27-item criterion-referenced test focuses on a broad range of astronomy content knowledge	Mainly undergraduate	Slater (2014)
The Astronomy and Space Science Concept Inventory (ASSCI)	211-item test aimed at measuring a broad range of astronomy and space science topics	School	Sadler et al. (2009)
Lunar Phases Concept Inventory (LPCI)	14-item test to measure conceptual change in students' knowledge of lunar phases after implementing a modified learning intervention	Undergraduate	Lindell (2001)
Moon Phases Concept Inventory for Middle School (MPCI-MS)	19-item instrument designed to assess student understanding of the underlying mechanisms for lunar concepts	School	Chastenay and Riopel (2020)
Star Properties Concept Inventory (SPCI)	25-item test that focuses on star properties	Undergraduate	Bailey et al. (2012)
Astronomy Knowledge Questionnaire (AKQ)	36-item test focussing on broad range of astronomical concepts built by modifying existing astronomy concept inventories to suit the southern hemisphere and also in line with the content of the astronomy unit	School and undergraduate	Lazendic-Galloway et al. (2017)
Conceptual Frameworks in Astronomy	23-item test based around a categorisation of five conceptual frameworks in astronomy: pre-science, geocentric, heliocentric, side-real/stellar, and galactic	Undergraduate	Finegold and Pundak (1991), and Pundak (2016)
Size Scale and Structure (S3CI)	24-item test that assesses students' knowledge of size, scale and structure in the context of various astronomy topics	Undergraduate	Gingrich et al. (2015) and Ladd et al. (2015)
Planet Formation Concept Inventory (PFCI)	20-item test to assess student learning on the topic of planet formation	Undergraduate	Simon et al. (2019)

Using the above reasoning challenges, we were able to construct a preliminary progression scale based on the SOLO taxonomy (Collis & Biggs, 1979), which combined with the quality of conceptions, provided a potential basis for a concept inventory. This current paper describes the development and validation of a concept inventory for cosmology in high school. The paper begins by introducing the research aim, then an explanation of the framing underpinning the study. Next, the paper highlights the methodological approach to the study, the results, and subsequent analysis. The paper concludes with a brief discussion of the findings, implications of the concept inventory, and concluding remarks.

Research Aim

Effective instruments to measure student understanding need to extend beyond declarative content knowledge or base-level comprehension to characterise the reasoning associated with deeper levels of conceptual knowledge. The aim of the research described in this paper is to develop and validate a concept inventory that can be used to appropriately characterise and monitor progression in student reasoning and learning in cosmology. The Cosmology Concept Inventory (CosmoCI) is a multi-dimensional concept inventory aimed at high school students, that moves beyond characterising lists of conceptions to enable exploration of progression in sophistication of student reasoning.

Framing This Study

Given this study aims to highlight the development and validation of a concept inventory for cosmology, we describe the framing of the study with regard to validity. The notion of validity is complex, and over many decades, there have been a range of debates about the various aspects of validity (e.g.: Cronbach, 1971; Cronbach & Meehl, 1955; Lissitz & Samuelsens, 2007; Messick, 1989; Sireci, 2007; Sireci & Parker, 2006). Traditionally, validity has been considered to be of three types: criterion, content, and construct. This has evolved into a unified theory of construct validity which has subsumed criterion and content validity as evidence for a more general framing of construct validity (Messick, 1989). What is considered valid depends on the context and aim and has to include expert judgement of the items in terms of their alignment with canonical/consensus ideas. Also relevant is the internal coherence across the levels, and alignment with students' thinking.

The American Educational Research Association, American Psychological Association, and National Council on Measurement, in the most recent version of the Standards for Educational and Psychological testing (2014), define validity as the “degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11).

From an educational perspective, Lissitz and Samuelsens (2007) emphasise the importance of content validity; however, Sireci (2007) states that “A serious effort to validate use of an educational test should involve both subjective analysis of test

content and empirical analysis of test score and item response data.” (p.481). Therefore, content validity on its own is not adequate. More recently, Sireci (2016) argues that the interpretation of test scores “is part of validation, and partly what validity refers to. Validating interpretations of test scores is a necessary component of any validation endeavour. However, it is not sufficient for defending the use of a test for a particular purpose” (p. 231).

This study aims to develop a concept inventory tool that will tap into students’ knowledge and reasoning in cosmology concepts, aligned with the four dimensions of cosmology developed by Salimpour et al. (2020b): size and scale, spacetime location, composition of the universe, and evolution of the universe. This tool is intended to.

- act as a pre-test to explore students’ knowledge in relation to key concepts in cosmology
- alert teachers to the key ideas in cosmology and provide them with an understanding of student thinking and reasoning
- provide a stimulus for classroom discussion
- provide a tool to monitor student learning

Methodology

The methodological approach in this study consists of two parts; the first is the process undertaken for developing the cosmology concept inventory (CosmoCI). The second involves using Rasch analysis in the form of a partial credit model (PCM) (Masters, 1982) to establish a scale, and evaluate the internal coherence and utility of the instrument.

Development of CosmoCI

The development of CosmoCI is underpinned by a Design-Based Research (DBR) framework (Anderson & Shattuck, 2012; Collins, 1992; Collins et al., 2004). The DBR cycle used in this study is visualised in Fig. 1.

The first iteration of CosmoCI consisted of 23 open-ended and five multiple choice questions. The questions were developed by reviewing curriculum statements related to cosmology in 52 curricula, which covered the OECD countries, China, and South Africa (Salimpour et al., 2020a) and the International Baccalaureate (IB) Diploma programme, and using the curriculum statements to extract key concepts in cosmology that were prevalent across most curricula. These were then categorised into four overarching conceptual dimensions of cosmology: size and scale, spacetime location, composition of the universe, and evolution of the universe (Salimpour et al., 2020b). The questions drew on previous research carried out at an undergraduate level (Wallace, 2011), filtered through the research team’s experience in both cosmology and teaching cosmology. Mostly open-ended questions were used to allow students to express their reasoning without being restricted. The first iteration of CosmoCI was implemented with a pilot group of students ($n=75$). The analysis

Fig. 1 Visualisation showing the different aspect of the DBR cycle employed in this study. As is noted, each step is iterative, allowing insights to be used in the next step

provided preliminary insights into students' knowledge and reasoning in relation to concepts across the four dimensions and allowed for the preliminary development of a universal rubric for characterising student responses (Salimpour et al., 2020b). On this basis, also, the questions were refined.

The second iteration used a slightly refined version in the wording of questions based on student responses. This is because student responses for some questions hinted that (a) some students may not have understood the question and (b) students were not sufficiently prompted to explain their reasoning. In addition, for the five multiple-choice questions, students were now asked to explain the reasoning for their choice. This second iteration was implemented with a larger group of participants ($n=286$). The analysis of the student responses allowed for the refinement of the grading rubric to include finer level distinctions (initially 16 levels and then 14 levels) that attempted to capture the levels of student reasoning, while encompassing the variety of responses (Salimpour et al., 2021b, in review). The underlying framework for building the levels of student responses was based on the SOLO taxonomy

(Collis & Biggs, 1979). This was a natural outcome that had developed during the first round of analysis, except finer levels within each SOLO level were included to capture the range of student responses. The finer-grained analysis, although helpful from a research perspective, we felt would be challenging for teachers to use to gain a big picture view of student reasoning at various levels of progression. Further, the patterns of responses across the questions and dimensions were extremely complex to analyse. Therefore, the finer levels were eventually collapsed into four of the five SOLO levels: relational, multi-structural, uni-structural, and pre-structural (Salimpour et al., 2021b, in review). This approach allowed broader patterns of student reasoning to be extracted, mapped to the SOLO levels. The reason the fifth SOLO level — extended abstract — was not included is that the nature of the questions being asked and the type of responses being canvassed did not prompt this extended form of reasoning.

The concept inventory idea is essentially based on the distractor-driven multiple-choice test (DDMC) (Sadler et al., 2009). The use of a multiple-choice format allows student understanding to be objectively assessed because it is practically efficient for teachers to use in the classroom environment. The task in this study was to take the variety of student responses for each question to construct multiple choice options that were exemplars of the responses at each of the four SOLO levels. That is, for each question, each option represented reasoning at a different SOLO level, constructed through thematic analyses of the open responses. Several cycles of refinement to the wording were made by the research team. The options were carefully designed not to be identifiable as higher level based on the use of abstracted disciplinary language and/or the length of the option. The mapping to SOLO levels provides teachers with a framework to characterise and appreciate the level of sophistication in student reasoning. An example of this categorisation approach is shown in Table 2

Brief Overview of CosmoCI

The question structure of CosmoCI (Appendix) is based on four overarching conceptual dimensions (Fig. 2) each focussing on a particular aspect of Cosmology. Each dimension encompasses key concepts and discoveries in cosmology, for example the large-scale structure, dark energy, dark matter, and the cosmic microwave background, all of which are present in curricula and found in high school physics textbooks. CosmoCI encompasses both declarative knowledge and higher-level conceptual knowledge, placing the progression of student reasoning at the core of its structure. The content coverage is organised under the four overarching, fundamental dimensions described above (Fig. 2).

To illustrate the scope of CosmoCI and how the multiple-choice options are aligned to the SOLO taxonomy, Table 3 provides an example of a question from each dimension. It can be noted that at each SOLO level fundamental reasoning similarities can be seen which are explained in the “Essence of levels” as shown in Table 2.

Table 2 Example of the categorisation of student responses into the four SOLO levels and multiple-choice option constructed based on student responses. Evo3 is the question numbering for the analysis. Evo3 denotes the third question in the dimension Evolution of Universe

Question 23 (Evo3): In the context of astronomy, what is the Big Bang Theory?

SOLO level	Essence of levels	Examples of student open responses	Multiple choice option
Relational	At this level students typically have a reasonable grasp of disciplinary knowledge. More importantly students are typically familiar with disciplinary knowledge at a conceptual level that supports generalised principles and are able to reason using that knowledge	<i>It is the theory that there was an expansion of space that created the universe</i> <i>The Big Bang is the expansion of the Universe</i> <i>It is stages that created our Universe</i>	It is a theory for understanding how the universe has changed during various stages of expansion
Multi-structural	At this level students typically use multiple lines of reasoning, linking ideas together, although there can be limitations in accuracy or completeness. Students possess some disciplinary knowledge although it is not always coherent	<i>The Big Bang Theory is the idea of the creation of the universe</i> <i>The Big Bang Theory is the leading explanation about how the universe began. At its simplest, it says the universe as we know it started with a small singularity</i> <i>The big bang is a big explosion that created stars and planets and has created life on earth and maybe on other planets</i> <i>The Big Bang Theory is the theory that our Solar system was made with the crash that rapidly formed more atoms and later on the universe</i> <i>It's nonsense!!!</i> <i>The big Bang theory is a creation story exploring how the universe was made to this day</i>	It is a theory for the origin/creation of the universe, that proves an explosion of a tiny singularity led to the formation of the universe
Uni-structural	At this level, students typically use a single line of reasoning. Students are typically more inclined to draw on single facts, often alternative ideas or partly on intuition		It is a theory for how the Sun, planets, and life emerged spectacularly in the beginning
Pre-structural	At this level, students typically demonstrate alternative conceptions that are drawn solely from intuition, and not representing valid reasoning. This can manifest in three forms; denial, guessing and restating the question		I disagree with the Big Bang Theory. It is just a theory at this point and does not necessarily explain observable phenomena in the universe
I do not know		IDK Don't know No idea	I do not know

Size and Scale	Spacetime Location
The questions in this dimension concern the spatial relations of size and scale in the context of the Universe on the largest scales and the objects which it encompasses.	The questions in this dimension concern the spatial and temporal relations that underpin positioning or reference frames from the perspective of the Solar System and the Universe at large.
The questions in this dimension concern the fundamental composition of the Universe and how scientists have come to characterise this composition.	The questions in this dimension concern the evolutionary history of the Universe and the key concepts underpinning our understanding of this timeline.
Composition of the Universe	Evolution of the Universe

Fig. 2 Characterisation of the four conceptual dimensions of cosmology

Validation of CosmoCI

This third iteration of CosmoCI (Appendix) was implemented with a cohort of students in Australia and Sweden. The participants included high school students in grades 10 and 11 ($n=234$). The sample selection was based on random opportunistic sampling (Newby, 2014). Australia and Sweden are used as instances, given that curriculum statements related to cosmology are relatively homogeneous across curricula (Salimpour et al., 2020a), and the authors have knowledge of the curriculum and access to schools in Australia and Sweden.

This study uses an item response theory (IRT) approach to validation. IRT is an item level approach and in essence aims to explore the relationship between test item (question) difficulty, and student ability (e.g.: Boone et al., 2014; Hambleton & Jones, 1993). Basically, easier questions are accessible by most students, and higher ability students are more likely to answer correctly difficult questions compared to a lower ability student. One aspect of IRT models is that results are independent of the participant group. IRT and its associated models (1-, 2-, 3-parameter) are independent of the participants taking the test. It should be emphasised that although the Rasch Model in some instances is referred to as being a type of IRT (specifically 1-parameter IRT model), there is a fundamental philosophical difference "... in that one model, IRT, is altered to fit data and one model, Rasch, is not altered to fit data" (Boone et al., 2014, p. 453). This current study uses Masters' (1982) partial credit model (PCM), which is a development of the basic Rasch model, and shares the defining characteristics associated with other Rasch models. The PCM is used for polytomously scored items, which is to say that the distractors in a question are given different levels of credit because they demonstrate a particular level of knowledge or reasoning (Masters, 1988). Therefore, multiple-choice questions are not solely marked as "correct" or "incorrect." For tests designed for the PCM model, the reasoning underpinning every multiple-choice option is distinct, with responses providing teachers with insights into the fundamental challenges that students face

Table 3 Example of a question for each dimension

Dimension	Size and scale	Spacetime location	Composition	Evolution
Question	What is meant by the statement: "The universe is expanding on a large scale"	An astronomer in a distant galaxy observes and measures the velocities of many galaxies, what will the astronomer observe?	What is dark energy? How do scientists know about it?	What is the cosmic microwave background radiation and what is its significance?
Relational	It means clusters of galaxies are moving apart due to space being created between them	That the distant galaxies are moving faster than the closer galaxies, because of the expansion of the universe	It is energy believed to be causing the Universe to expand faster. Scientists know about it by studying a particular type of exploding star	The glow left over from the early universe. It is significant because it supports our theory for the evolution of the universe
Multi-structural	It means the universe is expanding as space is being created at an accelerated rate according to Hubble's law (now called Hubble-Lemaître law)	That galaxies are moving away at mostly the same speed, because of the uniform expansion of the universe	It is a type of energy that acts the opposite to gravity, like negative gravity. Scientists know about it because of the expanding universe and the Big Bang	It is the glow from the origin of the Big Bang. It is significant because it is proof that the Big Bang happened
Uni-structural	It means the Big Bang has caused the distance between the Milky Way and other galaxies to increase	That the movement of galaxies is quite random, because they have different gravitational pull	It is negative energy that causes galaxies to expand. Scientists know about it from studying supermassive black holes at the centre of galaxies	It is radiation that exists in space. It is significant because it is harmful to astronauts in space
Pre-structural	It means as time goes by, the universe expands on an extremely huge scale in all directions	That galaxies' speed depends mainly on their size and colour, because if they have more older red stars, they are larger and so move slower	It is a type of energy that is not useable, or detectable. Scientists know about it because of scientific theories and computer simulations	It is microwave radiation that is emitted by various objects in the universe. It is significant because if we did not have an atmosphere, it could harm us

with regards to complex concepts (see for example, Briggs et al., 2006). This principle of identifying student reasoning is at the heart of this study, and thus using the PCM offers a framework for pursuing this line of inquiry.

The raw data from the CosmoCI was run through a custom Python script to clean the data, and format it for use with the QUEST software package (Adams & Khoo, 1993). QUEST is an analysis package that implements Rasch analysis on both dichotomous and polytomous data.

Results and Analysis

The statistical analysis reveals a reasonable fit to the Rasch model within the acceptable range set by literature (Adams & Khoo, 1993). Figure 3 shows that the item scores fit well within the “tram lines” (dotted lines).

One of the key outputs from the analysis is a set of Wright Maps (Wilson, 2005). Wright Maps allow researchers to quickly see the relationship between item difficulty and student ability, by placing both on the same measurement scale (Logit scale – log of the odds) (Fig. 4). This scale is a probabilistic measure and not an actual measure. Wright Maps were generated from the data analysis for each of the four conceptual dimensions in cosmology: size and scale, spacetime location, composition of the universe, and evolution of the universe (Salimpour et al., 2020b). The right side of the Wright Map lists the

Fig. 3 Analysis of the item fit to the Rasch model. Dotted lines represent the boundaries of the model fit

Fig. 4 Wright map showing all the questions and choice options. Each X represents 3 students. The variation in font thickness differentiates between the four SOLO levels

items, with the notation used being a three-letter word (Siz, Loc, Com, Evo), followed by 'x.y', where x is question number in that dimension and y is the multiple-choice level. Therefore, Siz1.2, refers to question 1, multiple-choice level 2 option (uni-structural) in the dimension size and scale. The order of items increases in difficulty (bottom to top). The left side of the Wright Map lists the students represented by X (which can be any number of students), in order of increasing ability (bottom to top). A student located in the same line as an item indicates that the student has a 50% chance of choosing that item option. Items above the student's position are harder for that student (students have less than 50% chance of choosing that reasoning option), and items below that student's position are easier (students have greater than 50% chance of choosing that reasoning option).

Looking at the Wright Map (Fig. 4), it can be seen that the multiple-choice options for each question line up in the predicted SOLO reasoning order, providing a validation of the identification of the levels of reasoning underpinning the different choices in each question. There are, however, some outliers where the SOLO levels are inconsistent across questions which will be discussed below. Overall, the Wright Map also reveals that the multiple-choice options satisfactorily align with the ability of the students; i.e. the highest ability students are able to pick the multiple-choice options which are at a relational level (4).

There are some outliers in the choice options. For example, items Evo1.3 and Evo3.3 were both assigned to a multi-structural level in design of the instrument, yet Fig. 4 shows that they are among the least difficult items when implemented with students. Item Evo1.3 gives the age of the universe as “more than 100 billion years,” and item Evo3.3 states that the Big Bang Theory “is a theory for the origin/creation of the Universe, that proves an explosion of a tiny singularity led to the formation of the Universe.” The solution to this anomaly lies in the notion of student alternative conceptions. With regard to item Evo1.3, previous work shows that students reason that because the universe is so large, it must be extremely old (Salimpour et al., 2021b, in review). While for item Evo3.3, there is a prevalent alternative conception that the Big Bang is an explosion a point in space; this has been shown in other studies as well (Aretz et al., 2016; Wallace, 2011), and is perhaps owing to the way representations depict the Big Bang Theory (Salimpour et al., 2021a). Perhaps, it is not surprising that students find this option attractive due to its pervasiveness as a popular metaphor, distinct from choosing the response through the reasoning process it seems to represent. The reason these items are deemed multi-structural lies in the fact that students need to bring together different lines of reasoning which are sophisticated, albeit alternative.

In addition to the above, the outliers are also an indication of a more fundamental challenge when collapsing the SOLO levels. The fine-grained expanded (14 and 16) SOLO categories made distinctions that took into account the following aspects present in student responses:

- the sophistication of reasoning;
- whether students’ ideas were scientifically correct;
- whether students were expressing alternative conceptions in a considered way;
- the amount of detail/justification students providing in their responses; and
- whether the multiple-choice component (if present) of a question was correct.

Some of these distinctions were lost in collapsing the data. However, this was justified since the complexity of the open responses represented by those categories masked the broader patterns of reasoning seen more clearly when the number of categories was reduced. We argue that the simplification offered by SOLO allows the instrument to capture patterns in students’ levels of thinking across questions and the four conceptual dimensions (size and scale, spacetime location, composition and evolution of the universe). Nevertheless, we acknowledge that the SOLO levels cannot capture all elements of students’ reasoning (Biggs & Collis, 1989; Watson et al., 1995).

Discussion

This study highlighted the process of developing and validating a cosmology concept inventory — CosmoCI — targeted at the high school level. The Rasch analysis in the form of a PCM shows a good alignment of the multiple-choice items in terms of increasing levels of sophistication and difficulty in line with student ability. For every question, the levels line up in the predicted order. The iterative coding of student responses naturally fit the SOLO taxonomy which provided a general framework for characterising the level of sophistication of the reasoning underpinning student responses. One of the key challenges of this study was unpicking knowledge and reasoning, and their interaction in framing student responses. The use of two iterations of open-ended questions allowed relations between “level of knowledge” and “pattern of thinking” to be explored (Salimpour et al., 2020b) and subsequently coordinated in CosmoCI.

The complexities in student responses at first warranted a fine-grained approach and so the SOLO taxonomy was expanded to capture these; however, to capture broad parameters of reasoning collapsing to the four SOLO levels provided a suitable characterisation. While the SOLO taxonomy theoretically represents distinct levels, in practice, these have blurred interfaces that allow characteristics from one level to manifest in adjoining levels. This is particularly evident in the clustering of multiple-choice options in the Wright map (Fig. 4). For example, among the level 4 options, there are some level 3 options with the same Logit score. Nevertheless, the consistencies in ordering of levels in the map demonstrate that the SOLO taxonomy as used in this study provides a valid and useful guide to anchor the progression in sophistication of student reasoning and knowledge in cosmology.

CosmoCI, through the situating of responses in a progression of reasoning, can be used as an assessment tool pre- and post-instruction. In addition, representation of the narrative of how scientists have come to know that the dark energy makes up the majority of the universe, that the universe is 13.8 billion years old, or that the universe is undergoing accelerated expansion, coupled with the innate interest it piques in students, means the CosmoCI can be a powerful stimulus to opening up rich discussions in the classroom. This latter aspect of instigating discussions was supported by the feedback from teachers who implemented CosmoCI in the classroom. The collection of questions can set the stage for learning to begin and support teachers to frame learning activities, as much as it can evaluate student conceptual understanding and reasoning in cosmology. This idea of determining what the student knows, so that teaching can be framed accordingly, echoes the views of Ausubel (1968).

It could be argued that sustaining such discussions in the classroom by teachers who may not be confident with the subject area could prove

challenging. However, with support, teachers could extend discussions of the questions raised by CosmoCI to consider the epistemic practices of cosmology — the way that ideas are built on evidence. The discussion of ideas, views, theories, and hypotheses form a vital part of the epistemic practices of science. The questions in CosmoCI and their categorisation into the four conceptual dimensions of cosmology provide the scaffolding needed to focus and instigate such discussions. With that in mind, the authors of this study are in the process of developing a teaching sequence for cosmology that incorporates an extensive guide for teachers, which includes a guide on unpacking the CosmoCI instrument.

Given that CosmoCI is developed through a design-based research (DBR) approach, the next stage will be the refinement of CosmoCI over various classroom implementations and more structured feedback from teachers.

Conclusions

This study aimed to develop and validate a Cosmology Concept Inventory (CosmoCI) for high school that can be used to monitor progression in reasoning and learning in cosmology. The validation process, through the application of a partial credit model IRT, shows that it is possible to capture the reasoning levels of students in cosmology using a multiple-choice instrument. The SOLO taxonomy proved versatile in capturing levels of reasoning associated with conceptions in cosmology. The instrument has an educative purpose through this reasoning focus, allowing teachers to discern the types of reasoning associated with cosmological declarative knowledge and concepts, acting as a formative and summative assessment tool to capture student thinking. It can open up discussions for teachers and students, and the possibility of informed support for moving students' thinking in cosmology forward. The inclusion within the instrument of aspects of the evidence bases for cosmological knowledge also aligns it with current thinking about the need to represent scientific practices in teaching and learning science, and with increasing attention to epistemic knowledge as an appropriate outcome. A beneficial outcome from implementing CosmoCI in the first author's classroom and those of other teachers who were part of the study was the rich discussions instigated between students and the teacher, providing insights into their thinking and the opportunity to frame their learning.

Appendix

Cosmology Concept Inventory (CosmoCI).

Table 4

Table 4 The 28 questions for CosmoCI, it should be noted that the order of multiple-choice options for each question and the order of the questions were randomised for each student

Level	Size and scale
	Q1: Is the universe infinite?
4	Partially, because while there is a limit to how far we can see, the universe is continuously expanding
3	Yes, because it is continuously expanding, and will continue to expand forever
2	Yes, because we have as yet not discovered the limits of the universe
1	No, because it is expanding and there is limit to the edge of the expansion
	Q2: Does the universe have an edge?
4	The universe has an edge beyond which we cannot observe anything, but it may not have the sort of edge we are used to imagining
3	No edge has been discovered, because the universe is so huge, it is predicted to be close to 100 billion light years across. Even if it does it cannot be reached
2	Yes, because if something is expanding, it has an outer edge like a balloon being blown up. However, it cannot be reached
1	No, because the universe is infinite and expanding according to the Hubble law (now called Hubble-Lemaître law)
	Q3: Two ideas about the size and scale of the universe is that it is infinite (goes on without any limit or an edge), or finite (has an edge). Which of the below is the best example of something that is finite (limited) but unbound (has no boundary)?
4	The surface of a doughnut
3	The Earth, which is a sphere, there is a certain amount of it, but we do not fall off the edge
2	Time goes on forever, but every human has a limited amount of time
1	The expanding universe because there is a limit to the size, but it is expanding forever
	Q4: What is meant by the statement: “The Universe is expanding on a large scale.”
4	It means clusters of galaxies are moving apart due to space being created between them
3	It means the universe is expanding as space is being created at an accelerated rate according to Hubble’s law (now called Hubble-Lemaître law)
2	It means the Big Bang has caused the distance between the Milky Way and other galaxies to increase
1	It means as time goes by, the universe expands on an extremely huge scale in all directions
	Q5: What is meant by the statement: “The universe, on a large scale, is the same at all locations and the same in all directions.”
4	Whichever direction we look the distribution of galaxy clusters is similar, and does not depend on your location
3	Wherever you are in the universe and whichever direction you point your telescope, it will look much the same
2	I disagree with the statement; the universe is different when you look at the night sky. There are different stars and galaxies

Table 4 (continued)

Level	Size and scale
1	The universe is big wherever you are in the world, you can still see huge galaxies and clusters of galaxies
	Q6: It is predicted that the Andromeda Galaxy and our own Milky Way Galaxy are on a collision course. How can this be the case if the universe is said to be expanding?
4	Because the expansion is not happening between single galaxies
3	Because the gravitational pull between them is stronger than expansion
2	Because each of the galaxies is expanding, and so they seem to be moving towards each other
1	Because the force from the Big Bang sent them moving towards each other on the same trajectory
	Q7: Despite there being a vast number of galaxies and stars in any direction you look, why is the night sky not extremely bright?
4	Because the universe has a finite age, the light beyond a certain distance has not yet reached us
3	Because light gets weaker/dimmer as it travels long distances, this is based on the Inverse Square Law
2	Because they are far away, and our eyes are not sensitive enough to see them
1	Because as the Earth rotates on its axis, when it gets dark one side is not facing the Sun, so there is a blockage of light
	Q8: Which of the below responses gives a larger number?
4	The average distance between stars divided by the diameter of an average star. Because stars are very small compared to the distance between them
3	The average distance between galaxies divided by the diameter of an average galaxy. Because galaxies are very big, but they are much further apart than stars
2	The average distance between galaxy clusters divided by the diameter of an average galaxy cluster. Because clusters of galaxies are extremely large, and so far apart that we can't even see them
1	This question cannot be answered because we do not have enough information
	Spacetime location
	Q9: In relation to the centre of the universe, where are we located?
4	There is no centre to the universe, any location can be considered the centre
3	We are not located at the centre of the universe, because we are located in one of the spiral arms of the Milky Way, which is not the centre
2	We know we are at the centre, because scientists have measured galaxies moving away from us, as the universe expands
1	We are located near the centre, being the third planet from the Sun and the Universe is expanding away from us
	Q10: An astronomer in a distant galaxy observes and measures the velocities of many galaxies, what will the astronomer observe?
4	That the distant galaxies are moving faster than the closer galaxies, because of the expansion of the universe
3	That galaxies are moving away at mostly the same speed, because of the uniform expansion of the universe
2	That the movement of galaxies is quite random, because they have different gravitational pull
1	That galaxies' speed depends mainly on their size and colour, because if they have more older red stars, they are larger and so move slower
	Q11: How would you best describe our location relative to the Sun?
4	In the zone, where life is possible because the temperatures are not too high for liquid water to exist

Table 4 (continued)

Level	Size and scale
3	Orbiting the Sun, at an approximate distance of 150 million kilometres
2	Third planet from the Sun, after Venus but before Mars
1	Pretty far away, because the Sun looks small. This varies depending on the time of year
	Q12: How would you best describe our location relative to the centre of our galaxy, The Milky Way?
4	We are located in one of the spiral arms of the galaxy, not quite halfway between the centre and edge
3	We are located somewhere in the galaxy but not the centre, because there are other objects in the centre
2	We are the third planet from the Sun, located in the solar system, orbiting close to the centre of the galaxy
1	We are located in the centre of the galaxy with everything rotating around the Sun
	Q13: How can astronomers determine our location in the Milky Way galaxy?
4	By mapping distances, speeds, and relative positions of a variety of astronomical objects in the Milky Way galaxy
3	By measuring distances, positions, and motions of stars to find the centre of the Milky Way galaxy
2	By measuring the brightness of different stars in the Milky Way galaxy using advanced satellites
1	By using satellites and telescopes to observe different objects like planets, and galaxies
	Q14: How can scientists determine the approximate age of the universe based on observations of many galaxies?
4	Scientists can use graphical and other representations relating the speed of galaxies to their distances
3	Scientists use the Big Bang Theory, in the form of the Hubble parameter
2	Scientists can use the speed galaxies are travelling to figure out how much the Universe has expanded since the Big Bang
1	By using technology and telescopes, to measure how fast the galaxies are moving, as galaxies age their speed changes
	Composition of the universe
	Q15: What was created in the early/initial stages of the universe's evolution?
4	Fundamental particles
3	Hydrogen and helium
2	Stars and galaxies
1	Planets and stars
	Q16: In terms of percentage, what makes up the majority of the content in the universe?
4	Dark energy
3	Empty space and dark matter
2	Fundamental particles like electrons
1	Atoms and molecules like hydrogen
	Q17: What is dark energy? How do scientists know about it?
4	It is energy believed to be causing the universe to expand faster. Scientists know about it by studying a particular type of exploding star
3	It is a type of energy that acts the opposite to gravity, like negative gravity. Scientists know about it because of the expanding universe and the Big Bang

Table 4 (continued)

Level	Size and scale
2	It is negative energy that causes galaxies to expand. Scientists know about it from studying super-massive black holes at the centre of galaxies
1	It is a type of energy that is not useable, or detectable. Scientists know about it because of scientific theories and computer simulations
	Q18: What is dark matter? How do scientists know about it?
4	It is a type of matter that does not emit any electromagnetic radiation. Scientists detect its effects through gravity
3	Dark matter is a form of matter that is found more than normal matter in the universe. Scientists know about it from its effects on other objects in the Universe
2	Dark matter is matter that is not detectable. Scientists know about it because of research
1	Matter in the universe that is dark, like black holes. Scientists detected this by studying the region around supermassive black holes
	Q19: What are the ways that the chemical elements in the universe form?
4	Light elements were formed during the early stages of the universe; heavy elements are formed in stars and supernova explosions
3	Various elements are produced in the life cycle of stars and in the Big Bang
2	By nuclear fusion in the cores of stars, and also in explosion and collisions of stars
1	They form at the centre of planets, through chemical reactions, and by smashing particles together at high temperatures
	Q20: How has the temperature of the universe changed over its lifetime?
4	It has decreased/become colder due to the expansion of the universe
3	It has increased/become warmer because the Sun is getting larger and older
2	It has not changed/remained the same, because space has always been cold
1	It has increased/become warmer because of the nature of universal climate change
	Evolution of the universe
	Q21: What is the approximate age of the universe?
4	Tens of billions of years
3	More than 100 billion years
2	100 million to 1 billion years
1	Less than 100 thousand years
	Q22: What are the most important stages in the evolution of the universe?
4	Inflation > fundamental particles > light elements > stars > galaxies > planets
3	Big Bang > fundamental particles > hydrogen > galaxies > stars > planets
2	Fundamental particles > elements > dust > stars > planets
1	Genesis > Big Bang > stars > galaxies > Sun > Earth
	Q23: In the context of astronomy, what is the Big Bang Theory?
4	It is a theory for understanding how the universe has changed during various stages of expansion
3	It is a theory for the origin/creation of the universe, that proves an explosion of a tiny singularity led to the formation of the universe
2	It is a theory for how the Sun, planets, and life emerged spectacularly in the beginning
1	I disagree with the Big Bang Theory. It is just a theory at this point and does not necessarily explain observable phenomena in the universe
	Q24: What is the most convincing evidence used in support of the Big Bang Theory?
4	The cosmic microwave background, together with the expansion of space

Table 4 (continued)

Level	Size and scale
3	The chemical composition of the universe including fundamental particles, hydrogen, and helium
2	The Hubble law (now called the Hubble-Lemaître) which shows the expansion of galaxies. Observations show that galaxies are moving away from us
1	There is no evidence to support the Big Bang Theory; it is just a theory about the origin of the universe
	Q25: What existed before the Big Bang?
4	This question cannot be asked, as the universe is everything
3	Some form of energy, which is the absence of matter, just empty space
2	Some particles like hydrogen and/or young stars
1	Some power that is beyond the understanding of science, which we cannot comprehend
	Q26: What is the cosmic microwave background radiation and what is its significance?
4	The glow left over from the early universe. It is significant because it supports our theory for the universe
3	It is the glow from the origin of the Big Bang. It is significant because it is proof that the Big Bang happened
2	It is radiation that exists in space. It is significant because it is harmful to astronauts in space
1	It is microwave radiation that is emitted by various objects in the universe. It is significant because if we did not have an atmosphere, it could harm us
	Q27: The early universe is said to have gone through a period called the “Dark Ages.” What does this mean and how did it come to an end?
4	It is the era when stars had not formed yet, it ended when the first stars formed
3	It is the era when the universe was cool and dark; it ended when the Big Bang happened and the universe expanded
2	It is the era when dark matter and dark energy were forming; it ended when the first stars formed and exploded
1	It is the era which our current laws of physics cannot explain what we see from this evolutionary stage of the universe; it ended with the enlightenment
	Q28: Scientists say that the early universe went through a period of inflation. To what are they referring?
4	The period when universe underwent a huge expansion in a very short amount of time
3	The period when the universe expanded to a large scale of many billions of light years
2	The period of the Big Bang when the universe exploded to be the observable universe
1	The period when gases in the universe expanded to allow the first stars to form

Acknowledgements The authors are very grateful to all the teachers and students from schools in Australia and Sweden who participated in this study. The fruition of this project would not have been possible without your keen and enthusiastic support. The authors would like to thank the reviewers for their constructive feedback in preparing this manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Dr. Michael Fitzgerald is the recipient of an Australian Research Council Discovery Early Career Award (project number DE180100682) funded by the Australian Government.

Declarations

Ethics This project was approved by Faculty of Arts and Education Human Ethics Advisory Group (HEAG) under the terms of Deakin University's Human Research Ethics Committee (DUHREC): HAE-19-049.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, R. J., & Khoo, S.-T. (1993). *Quest: The interactive test analysis system*. Quest, Australian Council for Educational Research Press.
- American Educational Research Association. (2014). *Standards for educational and psychological testing, 2014 Edition*.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>
- Aretz, S., Borowski, A., & Schmeling, S. (2016). A fairytale creation or the beginning of everything: Students' pre-instructional conceptions about the Big Bang Theory. *Perspectives in Science*, 10, 46–58. <https://doi.org/10.1016/j.pisc.2016.08.003>
- Aretz, S., Borowski, A., & Schmeling, S. (2017). Development and evaluation of a construct map for the understanding of the expansion of the Universe. *Science Education Review Letters*, 2017, 1–8. <https://doi.org/10.18452/8216>
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt Rinehart and Winston.
- Bailey, J. M. (2009). Concept inventories for ASTRO 101. *The Physics Teacher*, 47(7), 439–441. <https://doi.org/10.1119/1.3225503>
- Bailey, J. M., Johnson, B., Prather, E. E., & Slater, T. F. (2012). Development and validation of the star properties concept inventory. *International Journal of Science Education*, 34(14), 2257–2286. <https://doi.org/10.1080/09500693.2011.589869>
- Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education*, 33(2), 151–163. <https://doi.org/10.1177/168781408903300205>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33–63. https://doi.org/10.1207/s15326977ea1101_2
- Chastenay, P., & Riopel, M. (2020). Development and validation of the moon phases concept inventory for middle school. *Physical Review Physics Education Research*, 16(2), 020107. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020107>
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Springer.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

- Collis, K. F., & Biggs, J. B. (1979). *Classroom examples of cognitive development phenomena: The SOLO taxonomy*. Education Department, University of Tasmania.
- Cronbach, L. J. (1971). Test validation. In *Educational measurement* (2nd ed., p. 443). American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- 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>
- Finegold, M., & Pundak, D. (1991). A study of change in students' conceptual frameworks in Astronomy. *Studies in Educational Evaluation*, 17(1), 151–166.
- Gingrich, E. C., Ladd, E. F., Nottis, K. E. K., Udomprasert, P., & Goodman, A. A. (2015). The Size, Scale, and Structure Concept Inventory (S3CI) for astronomy. In G. Schultz, S. Buxner, L. Shore, & J. Barnes (Eds.), *Celebrating science: Putting education best practices to work* (Vol. 500, pp. 269–279). Astronomical Society of the Pacific Conference Series. Retrieved from http://www.aspbos.org/a/volumes/article_details/?paper_id=37504
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hansson, L., & Redfors, A. (2006). Swedish upper secondary students' views of the origin and development of the universe. *Research in Science Education*, 36(4), 355–379. <https://doi.org/10.1007/s11165-005-9009-y>
- Kind, P., & Osborne, J. F. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sci.21251>
- Ladd, E. (Ned), Nottis, K., Udomprasert, P., & Goodman, A. A. (2015). Initial development of a concept inventory to assess size, scale, and structure in introductory astronomy. *US-China Education Review B*, 5(11), 689–700. <https://doi.org/10.17265/2161-6248/2015.11.001>
- Lazendic-Galloway, J., Fitzgerald, M., & McKinnon, D. H. (2017). Implementing a studio-based flipped classroom in a first year Astronomy course. *International Journal of Innovation in Science and Mathematics Education* (Formerly CAL-Laborate International), 24(5), 35–47. <https://openjournals.library.sydney.edu.au/index.php/CAL/article/view/10670>
- Lindell, R. S. (2001). *Enhancing college students' understanding of lunar phases* [Doctoral dissertation, University of Nebraska - Lincoln]. ETD Collection for University of Nebraska - Lincoln.
- Lindell, R. S., Peak, E., & Foster, T. M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. *AIP Conference Proceedings*, 883(1), 14–17. <https://doi.org/10.1063/1.2508680>
- Lissitz, R. W., & Samuelson, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437–448. <https://doi.org/10.3102/0013189X07311286>
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4), 279–297. https://doi.org/10.1207/s15324818ame0104_2
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>
- Miller, B. W., & Brewer, W. F. (2010). Misconceptions of astronomical distances. *International Journal of Science Education*, 32(12), 1549–1560. <https://doi.org/10.1080/09500690903144099>
- Newby, P. (2014). *Research methods for education* (2nd ed.). Routledge. <http://ebookcentral.proquest.com/lib/deakin/detail.action?docID=1734204>
- Nussbaum, J., & Novak, J. D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education*, 60(4), 535–550. <https://doi.org/10.1002/sci.3730600414>
- Offerdahl, E. G., Prather, E. E., & Slater, T. F. (2002). Students' pre-instructional beliefs and reasoning strategies about astrobiology concepts. *Astronomy Education Review*, 1(2), 5–27. <https://doi.org/10.3847/AER2002002>
- Prather, E. E., Slater, T. F., & Offerdahl, E. G. (2002). Hints of a fundamental misconception in cosmology. *Astronomy Education Review*, 1(2), 28–34. <https://doi.org/10.3847/AER2002003>
- Pundak, D. (2016). Evaluation of conceptual frameworks in Astronomy. *Problems of Education in the 21st Century*, 69(1), 57–75.

- Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2009). 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), 010111. <https://doi.org/10.3847/AER2009024>
- Salimpour, S., Bartlett, S., Fitzgerald, M. T., McKinnon, D. H., Cutts, K. R., James, ... Ortiz-Gil, A. (2020a). The gateway science: A review of astronomy in the OECD school curricula, including China and South Africa. *Research in Science Education*, 51(4), 975–996. <https://doi.org/10.1007/s11165-020-09922-0>
- Salimpour, S., Tytler, R., & Fitzgerald, M. T. (2020b). Exploring the cosmos: The challenge of identifying patterns and conceptual progressions from student survey responses in cosmology. In R. Tytler, P. White, J. Ferguson, & J. C. Clark (Eds.), *Methodological Approaches to STEM Education Research* (Vol. 1, pp. 203–228). Cambridge University Press.
- Salimpour, S., Tytler, R., Eriksson, U., & Fitzgerald, M. T. (2021a). Cosmos visualized: Development of a qualitative framework for analyzing representations in cosmology education. *Physical Review Physics Education Research*, 17(1), 013104. <https://doi.org/https://doi.org/10.1103/PhysRevPhysEducRes.17.013104>
- Salimpour, S., Tytler, R., Fitzgerald, M. T., & Eriksson, U. (2021b). Is the Universe infinite? Exploring high student conceptions of cosmology concepts using open-ended surveys. (In Review).
- Simon, M. N., Prather, E. E., Buxner, S. R., & Impey, C. D. (2019). The development and validation of the Planet Formation Concept Inventory. *International Journal of Science Education*, 41(17), 2448–2464. <https://doi.org/10.1080/09500693.2019.1685140>
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477–481.
- Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23(2), 226–235. <https://doi.org/10.1080/0969594X.2015.1072084>
- Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*, 25(3), 27–34. <https://doi.org/10.1111/j.1745-3992.2006.00065.x>
- Slater, E., Morris, J., & McKinnon, D. H. (2018). Astronomy alternative conceptions in pre-adolescent students in Western Australia. *International Journal of Science Education*, 40(17), 2158–2180. <https://doi.org/10.1080/09500693.2018.1522014>
- Slater, S. J. (2014). The development and validation of the test of astronomy standards (TOAST). *Journal of Astronomy & Earth Sciences Education (JAESE)*, 1(1), 1–22. <https://doi.org/10.19030/jaese.v1i1.9102>
- Slater, S. J., Tatge, C. B., Bretones, P. S., Slater, T. F., Schleigh, S. P., McKinnon, D. H., & Heyer, I. (2016). iSTAR first light: Characterizing astronomy education research dissertations in the iSTAR database. *Journal of Astronomy & Earth Sciences Education (JAESE)*, 3(2), 125–140. <https://doi.org/10.19030/jaese.v3i2.9845>
- Trouille, L. E., Coble, K., Cochran, G. L., Bailey, J. M., Camarillo, C. T., Nickerson, M. D., & Cominsky, L. R. (2013). Investigating student ideas about cosmology III: Big Bang Theory, expansion, age, and history of the Universe. *Astronomy Education Review*, 12(1), 010110. <https://doi.org/10.3847/AER2013016>
- Trundle, K. C., Atwood, R. K., & Christopher, J. E. (2007). Fourth-grade elementary students' conceptions of standards-based lunar concepts. *International Journal of Science Education*, 29(5), 595–616. <https://doi.org/10.1080/09500690600779932>
- 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](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
- Wallace, C. S. (2011). An investigation into introductory astronomy students' difficulties with cosmology, and the development, validation, and efficacy of a new suite of cosmology lecture-tutorials. *Astrophysical & Planetary Sciences Graduate Theses & Dissertations*. http://scholar.colorado.edu/astr_gradetds/9
- Wallace, C. S., & Bailey, J. M. (2010). Do concept inventories actually measure anything? *Astronomy Education Review*, 9(1), 010116. <https://doi.org/10.3847/AER2010024>
- Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation*, 1(3), 247–275. <https://doi.org/10.1080/1380361950010303>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Routledge.

Authors and Affiliations

Saeed Salimpour^{1,2,3,4} · **Russell Tytler**¹ · **Brian Doig**¹ ·
Michael T. Fitzgerald^{5,6} · **Urban Eriksson**⁷

¹ Deakin University, Victoria, Burwood, Australia

² Office of Astronomy for Education, International Astronomical Union, Heidelberg, Germany

³ Haus Der Astronomie, Heidelberg, Germany

⁴ Max Planck Institute for Astronomy, Heidelberg, Germany

⁵ Deakin University, Burwood, Victoria, Australia

⁶ Las Cumbres Observatory, Goleta, CA, USA

⁷ National Resource Centre for Physics Education, Department of Physics, Lund University, Lund, Sweden