Parker, Mark; Hedgeland, Holly; Braithwaite, Nicholas St.J. and Jordan, Sally E. (2024). GRCI: An investigation into the feasibility of a General Relativity Concept Inventory. European Journal of Science and Mathematics Education, 12(4) pp. 489–501.

European Journal of Science and Mathematics Education, 2024, **12**(4), 489-501
ISSN: 2301-251X (Online)

OPEN ACCESS

Research Article

GRCI: An investigation into the feasibility of a General Relativity Concept Inventory

Mark A. J. Parker [1*]
0000-0002-1984-9624

Holly Hedgeland [2]
0000-0003-3703-7942

Nicholas St. J. Braithwaite [3]
0000-0002-1586-3736

Sally E. Jordan [1]
0000-0003-0770-1443

[1] School of Physical Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UNITED KINGDOM
[2] Clare Hall, University of Cambridge, Herschel Road, Cambridge, CB3 9AL, UNITED KINGDOM
[3] Faculty of Science, Technology, Engineering & Mathematics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UNITED KINGDOM
[*] Corresponding author: mark.parker@open.ac.uk

Citation: Parker, M. A. J., Hedgeland, H., Braithwaite, N. St. J., & Jordan, S. E. (2024). GRCI: An investigation into the feasibility of a General Relativity Concept Inventory. *European Journal of Science and Mathematics Education, 12*(4), 489-501. https://doi.org/10.30935/scimath/15018

ARTICLE INFO

Received: 3 Jun 2024

Accepted: 25 Jul 2024

ABSTRACT

The study outlines the early-stage development of a free-response General Relativity Concept Inventory (GRCI), an educational instrument designed to test for conceptual understanding of General Relativity. Data were collected for the study by having 26 participants from General Relativity courses work through the questions on the GRCI. Interviews were conducted with four of the participants to gain further insight about their experience of working through the GRCI. The written responses revealed that participants were proficient when answering questions which required mathematical thought processes, but were more limited when answering questions which required conceptual and physical thought processes. The interviews revealed that participants found that free-response questions were appropriate to test for conceptual understanding of General Relativity. Participants identified that General Relativity has physical interpretations and mathematical constructs, and both are important to understand the theory. Participants thought that the GRCI could be given a formative purpose in a teaching context. The study was *proof of concept* in scope, with the aim of highlighting important points pertaining to the feasibility and development of the GRCI. Additional work to further investigate the above points highlighted by the study is encouraged.

Keywords: general relativity, concept inventories, student reaction, conceptual understanding, physics education

INTRODUCTION

Concept inventories are instruments used in Physics Education Research to measure the conceptual understanding of students (Smith & Tanner, 2010) and to investigate the effectiveness of various teaching interventions (Porter et al., 2014). This is achieved by getting the students to answer the questions on the concept inventory first before instruction, and then a second time after instruction. Comparison of the pre-instruction and post-instruction scores reveals which concepts students still struggle with after instruction

(Bailey et al., 2012). This allows teachers to develop new teaching approaches and methods to focus on teaching these concepts.

The first concept inventory was the Force Concept Inventory (FCI) (Hestenes et al., 1992), designed to test conceptual understanding of Newtonian mechanics. The FCI includes 30 multiple-choice questions, each with one correct answer and a number of incorrect distractors, reflecting common misconceptions. Following on from the FCI, several other concept inventories were developed for a variety of subject areas in physics and astronomy. Examples of these inventories include the Force and Motion Conceptual Evaluation (Thornton & Sokoloff, 1998), the Astronomy Diagnostics Test (Hufnagel, 2002; Zeilik, 2003), and the Star Properties Concept Inventory (Bailey et al., 2012).

Concept inventories are typically designed to contain minimal mathematical content (Lindell et al., 2007), owing to their focus on evaluating conceptual understanding. In addition, concept inventories are usually administered as a pre-test and a post-test, as previously noted (Bailey et al., 2012). At first glance, the *concept inventory* format seemingly appears unsuitable for mathematical physics topics, as students are unlikely to have sufficient prerequisite knowledge to answer the concept inventory questions before instruction has taken place. However, students are still capable of having intuitive (and often incorrect) ideas about mathematical physics topics before instruction, and these ideas should be modified somewhat after instruction has taken place. Following on from this idea, a number of concept inventories corresponding to mathematical physics topics have been developed and tested. For example, there are concept inventories covering content from Quantum Mechanics (Dick-Perez et al., 2016) and Electromagnetism (Baily et al., 2017; Ding et al., 2006). Furthermore, there is a concept inventory which tests understand of content from Special Relativity (Aslanides & Savage, 2013). Of note, this inventory does not cover topics from General Relativity.

General Relativity is a mathematical physics topics which is noted in the literature as being difficult to teach (Burko, 2017). Alongside the complex mathematics which underpins the subject, General Relativity also has a deep conceptual grounding, exemplified by the original thought experiments used by Einstein to conceive and develop the subject (Stannard et al., 2017). Because of the abstract nature of the subject, students will not have the same kind of pre-existing conceptions about General Relativity as they do with Newtonian Mechanics, where students already have an *Aristotelian* understanding of Force and Motion from their everyday experiences. However, General Relativity still appears in daily life, even if students do not realize this. For example, GPS systems and motions of celestial bodies are both underpinned by concepts from the subject. Further, students will have intuitive ideas about the nature of space and time from their own experiences, and many of these break down on the relativistic scale (Conlon et al., 2017). Consequently, General Relativity has some surprising and counter-intuitive results, such as black holes and the expansion of the Universe.

Much like the other cases from mathematical physics above, it is unlikely that students would have complete understanding about the nature of space and time without instruction on the subject. However, physics students typically have exposure to popularized concepts, such as curvature around massive objects, prior to studying General Relativity in detail. This gives value to investigating the suitability of a pre-test, post-test format for a concept inventory based on the topic. Putting these ideas together, developing a concept inventory for content from General Relativity could provide useful insight into the teaching, learning and understanding of the subject. This idea forms the rationale for the current study.

The aim of this work is to outline the early-stage development of a new concept inventory, the General Relativity Concept Inventory (GRCI), which covers topics from the subject of General Relativity. The questions on the GRCI are in the free-response format, meaning that students have to write their own responses to the questions. Previous work has been conducted with the aim of developing a free-response concept inventory for the subject of Newtonian mechanics (Parker et al., 2023; Rebello & Zollman, 2004), which is a less mathematically involved physics topic. Within this context, the current work acts as a *proof of concept* for the idea of employing free-response questions in a concept inventory for a mathematically involved physics topic, General Relativity. As such, this work details analysis of data gathered from responses to the GRCI questions and corresponding interviews with physics students who answered the GRCI questions.

The study was guided by the following research questions:

RQ1: How can a concept inventory containing free-response questions be used to test conceptual understanding of General Relativity?

RQ2: What thought processes are used by students to answer the GRCI questions, and how do students react to the conceptual nature of the questions posed in the GRCI?

METHODS

Background

Questions were required for the proposed GRCI in order to conduct the study. Material from two General Relativity courses at a UK university and consultation with academics whose backgrounds were in Astrophysics, Relativity and Cosmology were used to investigate which concepts are covered in a typical General Relativity course. Based on these sources, four topics were chosen as the basis for the GRCI. These topics are labelled as T1, T2, T3 and T4 below.

- Frames of Reference [T1].
- Ingredients for General Relativity (Principle of Equivalence and Curvature) [T2].
- The Einstein Field Equations [T3].
- General Relativity and Cosmology [T4].

The GRCI contained 10 questions, with two questions corresponding to [T1]; three questions corresponding to [T2]; two questions corresponding to [T3]; and three questions corresponding to [T4]. These were the GRCI questions used to conduct the study.

Data Collection

Response data was required to investigate and develop the GRCI questions. Since this work was *proof of concept* in nature, the GRCI questions were still in the development phase during the study. Taking this context into consideration, data were collected from students who had some familiarity with the subject of General Relativity. This choice of participants was made because students who knew some content from the subject matter would be better poised to identify if there were any issues with the questions than students who had not previously encountered the subject. As such, this choice was made to facilitate the development process.

The Institute of Physics (2024) does not require General Relativity to be taught in the physics degrees it accredits, so within the UK and Ireland, General Relativity is not taught at all universities, though it is sometimes included. Data collection efforts focused on students taking General Relativity courses at three very different UK universities. In what follows, the three institutions are referred to as Institution A, Institution B and Institution C.

Students from the three higher education institutions were invited to attempt the GRCI questions. Students who participated in the study were close to the end of an undergraduate physics degree. These students all had similar subject backgrounds and levels of study experience. Participants would have had different levels of previous exposure to the topics covered in the GRCI, based on what had been covered on each General Relativity course. No effort was made to gather students of different demographics, as investigating responses based on demographics was not the aim of this study. Further, no feedback was given to GRCI test-takers about their performance, as concept inventories do not usually provide feedback (Parker et al., 2022).

Each institution administered the GRCI in a different way. At Institution A, the GRCI questions were administered online. Nine students completed the GRCI in this way, and the results were collected by downloading them directly in electronic form. At Institution B, the GRCI questions were administered on paper; 13 students completed the GRCI in this way, and the results were manually collected together into a spreadsheet. At Institution C, the participants did the GRCI questions on paper, and a short interview about the experience followed. Four students completed the GRCI in this way, and each of them received an Amazon voucher worth

£20 in appreciation of their involvement. The written GRCI responses were manually collected together into a spreadsheet, and the interview responses were transcribed manually.

The study made use of 26 participants, and was *proof of concept* in scope. The data collected from participants were to give insight into the feasibility of the GRCI, including student reaction and written responses to the questions, and to highlight any development issues with the instrument. It is noted that testing and interviewing a greater number of participants would have been useful, but there were not large numbers of students available on General Relativity courses. This is acknowledged as a limitation of the study.

Data Analysis

There were two components to the data. The first were the written answers given by the 26 participants to the GRCI questions. The second were the verbal responses given by the four participants from Institution C to the interview questions. Although the study made use of a smaller number of participants, the data collected provided a rich data set to analyse. Both parts of the data were used to address different aspects of the research. The GRCI interview responses were used to find out what the participants thought of the GRCI and to gauge their reaction to it, while the written responses were used to investigate the concepts that students used when giving their answers. The findings from both parts of the data were used to draw conclusions about the feasibility and development process of the GRCI.

Following on from the above considerations, both parts of the data required different treatment. Owing to the variety of the data and the smaller sample size, a *Mixed Methods* (Catalyst Harvard, 2024) approach was selected to analyse the data. The Mixed Methods approach combines different research methods, drawing on the strengths and advantages of each. Mixed Methods provides a meaningful description and explanation of the available data, facilitating the interpretation of results and helping to understand the broad applicability of findings from studies with smaller sample sizes. This approach lends its strengths to the current *proof of concept* study, where the number of participants was small, while the data collected was rich in meaning.

The GRCI response data from the 26 participants formed the first data set. The responses were gathered into a spreadsheet, and marked against model answers. Marking was binary, with a mark of 1 awarded for an answer that was judged to be correct, and a mark of 0 awarded for an answer that was judged to be incorrect. The content and semantics of the responses were examined to investigate what concepts the participants were using to answer the GRCI questions.

The GRCI interview data formed another data set, and *Thematic Analysis* (Braun & Clarke, 2006; Braun et al., 2014) was used to find underlying themes. Thematic Analysis can be used to analyse a qualitative data set drawn from a study which uses a small number of participants. The data are difficult to understand in its raw form, so Thematic Analysis reduces the data into a form which can be interpreted. This reduction process results in the eponymous themes of the method. This approach prevents the investigator from drawing arbitrary conclusions from the data, since the themes are emergent from the rate of occurrence of underlying codes within the data set. The version of Thematic Analysis outlined by the University of Auckland (2024) was used to analyse the interview data in the current study.

RESULTS

Findings from the GRCI Responses

The concepts used in correct answers to the GRCI questions and misconceptions used in incorrect answers to the GRCI questions are considered below. The topic of *Frames of Reference* was tested in Q1 and Q2 of the GRCI. For Q1, correct answers referred to Newtonian gravitation being an approximation, whereas incorrect answers repeated information from the question without reaching the required conclusion. For Q2, correct answers referred to the observer being in a non-accelerating state of motion, while incorrect answers were unable to reach this conclusion about the observer’s state of motion.

The topic of *Ingredients for General Relativity (Principle of Equivalence and Curvature)* was tested in Q3, Q4 and Q5 of the GRCI. For Q3, correct answers referred to the observer being unable to tell whether their motion was due to the effects of gravity or accelerating motion. Incorrect answers to Q3 arose where students were unable to properly interpret the Principle of Equivalence. For Q4, correct answers referred to the Curvature of the non-Euclidean space; incorrect answers instead named a type of geometry, such as spherical geometry, without giving further explanation. For Q5, correct answers identified that Curvature gives rise to the non-

Euclidean spaces required for the geometric view of spacetime used in General Relativity, whereas incorrect answers failed to make this realization.

The topic of *The Einstein Field Equations* was tested in Q6 and Q7 of the GRCI. For Q6, correct answers noted that the Einstein Field Equations determine the geometry of the spacetime; incorrect answers failed to recognize this. For Q7, correct answers recognized that the energy-momentum tensor determines the matter-energy inventory of the spacetime. Incorrect answers to Q7 did not make this connection between energy, momentum and spacetime.

The topic of *General Relativity and Cosmology* was tested in Q8, Q9 and Q10 of the GRCI. For Q8, correct answers identified that applying the Cosmological Principle simplifies the situation by assuming that the Universe is homogeneous and isotropic, while incorrect answers were not able to capture this idea. For Q9, correct answers noted that the matter-energy density of the Universe is thought to be at the critical density, whereas incorrect answers did not recognize this. For Q10, correct answers noted that it is currently thought that the Universe will expand forever. On the other hand, incorrect answers to Q10 typically listed other possible scenarios for the fate of the Universe.

The written responses to the GRCI questions revealed strengths and weaknesses in the understanding of the students. The incorrect answers to Q1 and Q2 revealed that students struggled to understand *Frames of Reference*. This topic has been identified as being important for the overall understanding of General Relativity in the work of Semon et al. (2009), as it provides a historical and conceptual link between Special Relativity and General Relativity. The incorrect answers to Q3 showed that students were unable to apply the *Principle of Equivalence* to solve a problem. This is consistent with the findings of Bandyopadhyay and Kumar (2010), who highlighted the same issue.

Q4 and Q5 were about *Curvature*, whereas Q6 and Q7 were based on the *Einstein Field Equations*. Each of these questions tested General Relativity topics that were mathematical in nature. Q4, Q5, Q6 and Q7 were generally well-answered by the students. This outcome may be expected because General Relativity courses are often mathematically grounded (Hartle, 2008). Q8, Q9 and Q10 were based on *General Relativity and Cosmology*. The incorrect answers to these questions showed that students had difficulties interpreting the results of General Relativity in the Cosmology context. This finding agrees with the work of Conlon et al. (2017), which found that students have various ideas about the fate of the Universe.

The findings from the written responses to the GRCI questions can be summarized as follows. Incorrect answers to Q1, Q2 and Q3 showed that students found the physical interpretation of General Relativity to be difficult. Correct answers to Q4, Q5, Q6 and Q7 showed that students had a good understanding of the mathematical aspects of General Relativity. Incorrect answers to Q8, Q9 and Q10 showed that students struggled to apply the results of General Relativity to the Cosmology context. The students did well when answering questions based on mathematical ideas, but encountered difficulties when answering questions based on physical concepts. This may be a result of the structure of the General Relativity instruction that the students received. This finding could be a useful consideration for the development of concept inventories for other mathematical physics subjects.

Findings from the GRCI Interviews

For the interview data, *Thematic Analysis* identified 8 codes, which grouped together into 3 themes. These were "Free-response questions are appropriate to test for conceptual understanding of General Relativity", "General Relativity joins physical interpretations with mathematical constructs, and both are important when fluent with the theory", and "The GRCI can be given a formative purpose". Each of these themes are discussed below.

Findings related to the "Free-response questions are appropriate to test for conceptual understanding of General Relativity" theme

There were 2 codes associated with the "Free-response questions are appropriate to test for conceptual understanding of General Relativity" theme. This theme was coded 26 times overall. The codes associated with this theme are presented in **Table 1**. Unless otherwise stated, the occurrence of the codes was more or less equal between the four participants.

Table 1. Codes associated with the "Free-response questions are appropriate to test for conceptual understanding of General Relativity" theme

Code	Number of times coded
Free-response questions made participants think (C1)	21
Free-response questions test understanding more thoroughly than multiple-choice questions (C2)	5

Table 2. Codes associated with the "General Relativity joins physical interpretations with mathematical constructs, and both are important when fluent with the theory" theme

Code	Number of times coded
In responding to the GRCI questions, participants referred to the mathematical interpretations of General Relativity (C3)	5
In responding to the GRCI questions, participants referred to the physical interpretations of General Relativity (C4)	29
In responding to the GRCI questions, there were cases where the answer was difficult to articulate using words (C5)	6
It was recognized that General Relativity could be used in the context of science engagement (C6)	2

For code C1, the participants noted that the open-ended format of the free-response GRCI questions was different from the General Relativity questions that they were accustomed to, which typically sought algebraic solutions. Furthermore, they noted that the free-response question type made them think about their answers, since a semantic based approach was required to come up with the words to answer with.

In code C2, the participants contrasted the free-response format of the GRCI questions with their experience of multiple-choice questions. They noted that free-response questions prevent test-takers from making use of *eliminate and guess* and similar strategies when answering the questions. In addition, codes C1 and C2 highlighted that the participants were aware that the questions could be asked in multiple-choice format. This suggests that the participants might have had previous exposure to concept inventories or other related types of physics assessment.

Through codes C1 and C2, participants identified that free-response questions were appropriate for testing for conceptual understanding of General Relativity, since the open-ended format gave them the chance to express what they were thinking. Participants felt that having to come up with their own answers was better suited for testing their General Relativity understanding than multiple-choice questions. With free-response questions, participants noted that they had to think carefully about what they knew, and write answers to articulate their ideas. This experience was contrasted with the multiple-choice question format, where they would have selected from a list of options that somebody else had already written.

One participant went into further detail about how they approached different question types. They reflected that mathematical questions can be solved in a procedural manner; multiple-choice questions can be answered using an *eliminate and guess* strategy; and free-response questions require students to apply their subject knowledge to be answered. The above observations indicate that different types of questions can be used to test various types of understanding of General Relativity, such as mathematical understanding and conceptual understanding.

Findings related to the "General Relativity joins physical interpretations with mathematical constructs, and both are important when fluent with the theory" theme

There were 4 codes related to the "General Relativity joins physical interpretations with mathematical constructs, and both are important when fluent with the theory" theme. Overall, this theme was coded 40 times. The codes associated with this theme are given in **Table 2**. Code C6 was referred to exclusively by participant P1. It was retained for the analysis because it raised an interesting point about the appeal of General Relativity as a subject.

In code C3, participants discussed the mathematical interpretation of General Relativity. They noted that no calculations were required to answer the GRCI questions. For code C4, participants noted that they needed to understand what they were writing about in their answers, because they were thinking about the physical interpretations of the theory. In addition, participants identified the importance of understanding how the

Table 3. Codes associated with the "The GRCI can be given a formative purpose" theme

Code	Number of times coded
Participants used experience of doing GRCI to reflect upon their understanding of General Relativity (C7)	29
Within an educational context, participants identified that the GRCI could be used as a teaching tool (C8)	5

theory of General Relativity was originally formulated, along with its physical and philosophical implications. However, participants acknowledged that the physical interpretations of General Relativity could be difficult to articulate using words. This idea was captured through code C5.

Through codes C3 and C4, participants identified that General Relativity is a mathematical and abstract topic, and that it has physical consequences that are applicable in the everyday world. In code C5, participants found that it was difficult to express these mathematical and physical ideas in words. Code C6 relates these ideas to a science engagement context. One participant noted that General Relativity is an important topic which would be of interest to a general audience, but that it would be difficult to explain the nuances of the theory to such an audience. This is because a general audience would not have the mathematical fluency required to understand the meaning and consequences General Relativity. In addition, the theory is difficult to describe and explain using non-specialized language.

Through code C3, participants noticed that the GRCI questions were not mathematical, which contrasted with their previous experience of General Relativity questions. Code C4 showed that the questions on the GRCI made the participants look beyond the mathematics of the situations, and made them think about the scenarios from a physical viewpoint. These points were identified as being both challenging and rewarding tasks in codes C5 and C6. The GRCI questions helped the participants to understand the physical consequences of the theory, thereby giving it some context within a larger conceptual framework. Related to this, participants found that being tested on General Relativity conceptual understanding made them go back and consider how the theory was formulated, what they key concepts really meant, and how these fitted together to form the theory of General Relativity. These reflections showed that some participants thought of General Relativity in mostly mathematical terms, which highlighted weaknesses when it came to interpreting and understanding the results in a physical context. Taken together, the above points indicated that doing the GRCI questions was a useful educational exercise for the participants.

One participant considered attempting to use General Relativity in a scientific outreach or engagement context. General Relativity is an integral part of understanding space and time, but very few people ever gain any exposure to the subject. General Relativity is frequently not taught as part of the undergraduate physics curriculum (Hartle, 2008; Institute of Physics, 2024), meaning that even students who have chosen to study physics may not encounter the subject during their studies. As such, only those who specialize in General Relativity are ever likely to gain significant exposure to the subject. These considerations would need to be taken into account in order to effectively use General Relativity in science outreach or engagement efforts.

Findings related to the "The GRCI can be given a formative purpose" theme

The "The GRCI can be given a formative purpose" theme was coded 34 times overall. This theme consists of 2 codes, and these are given in **Table 3**. In code C7, the participants talked through the answers they gave to the GRCI questions. Each participant reacted to how well they had done, and noted that they had not answered some of the questions well. Furthermore, all of the participants identified that they would have liked to get more detailed feedback on their work. The participants also discussed how the GRCI could be used for teaching purposes in a General Relativity course, which was captured in code C8.

C7 showed that the participants reflected on how well they had answered the questions on the GRCI. Some of the participants felt that they had forgotten the content required to answer the questions. This was because the topics covered by the GRCI had been studied several months previously. As such, there were instances where the participants were not confident in their answers. Furthermore, participants talked through their line of reasoning when discussing questions that they had found to be difficult. All participants felt as if it would have been useful to get feedback after completing the GRCI questions. Participants wanted to know whether their answers were correct, and to highlight gaps in their understanding of the material. These

Table 4. Different levels of the 'knowledge' dimension of Bloom's Revised Taxonomy

Knowledge level	Meaning
Factual knowledge	The student can recall basic facts related to the subject.
Conceptual knowledge	The student is able to put different elements of the subject into a knowledge structure.
Procedural knowledge	The student can apply methods to solve problems within the subject.
Metacognitive knowledge	The student is self-aware of their own levels of knowledge and cognition.

Table 5. Different levels of the 'cognitive process' dimension of Bloom's Revised Taxonomy

Cognitive process level	Meaning
Remember	The student can recall basic elements related to the subject.
Understand	Student is able to determine meaning from recalled elements related to the subject.
Apply	The student can carry out a method or procedure to solve a problem related to the subject.
Analyse	The student can see how different parts of subject fit together as part of a bigger picture view.
Evaluate	The student makes judgements based on their own understanding of the subject matter.
Create	The student puts the different pieces of the subject together to form a coherent worldview.

findings agree with the idea from the literature that students in general like to get feedback on their work (Parker et al., 2022; Zhu et al., 2020).

Looking forward, some of the participants went on to discuss potential future uses for the GRCI. These ideas were captured by code C8. One participant suggested the possibility of using the GRCI in a small group setting to address weaknesses and build conceptual understanding. This would use the GRCI as a teaching intervention with real-time feedback, which is a different function from that of a standard concept inventory. The suggestion of this participant is linked to the idea that feedback could be used to facilitate with conceptual understanding, which has been suggested in the literature (Bulut et al., 2019).

DISCUSSION

Discussion – Learning General Relativity

Bloom's Taxonomy can be applied to understand the above ideas about gathering and building General Relativity subject knowledge. Bloom's Taxonomy is a classification system for educational objectives. The taxonomy was originally proposed as a one-dimensional construct (Bloom, 1956) based on knowledge. This initial construct was subsequently revised into Bloom's Revised Taxonomy (Krathwohl, 2002), which includes the two dimensions of knowledge and cognitive process. The various tiers of these dimensions are given in **Table 4** and **Table 5**.

Within the context of Bloom's Revised Taxonomy, the importance of conceptual understanding for the mastery of General Relativity is highlighted through the *Conceptual Knowledge* level from **Table 4** and the *Understand* level from **Table 5**. Conceptual understanding of General Relativity is underpinned by having mathematical competence with the formulae and techniques of the subject, as well as having the ability to interpret results in a physical setting. Referring again to the Bloom's Revised Taxonomy, the mathematical competences required to solve General Relativity problems are underpinned by the *Procedural Knowledge* level from **Table 4** and the *Apply* level from **Table 5**. The capacity to interpret the results of such calculations in a physical context pertains to the *Analyse* and *Evaluate* levels from **Table 5**.

Students with lower levels of expertise may struggle with one or both of the mathematical and physical aspects of General Relativity, which in turn hinders their conceptual understanding of the subject. However, Bloom's Revised Taxonomy is a continuum, which means that such students have the potential to improve their conceptual understanding by studying General Relativity and mastering both the mathematical and physical aspects of the subject. This level of mastery comes when the student gains the ability to independently move between both interpretations of the subject. Furthermore, this subject mastery allows the student to explain the mathematical and physical interpretations to peers in their own words.

Within the context of the current GRCI study, the participants noted that General Relativity is a highly mathematical subject, and that they had been taught it in a way that reflected this. In addition, participants contrasted the mathematical aspects of the subject with the way that General Relativity can be applied physically in the everyday world. This illustrated that the participants were aware that both mathematical and physical interpretations were important for understanding of the theory. This indicates that the participants had started to work through the earlier levels of *cognitive process* outlined in Bloom's Revised Taxonomy. That said, the participants did not yet have a full understanding of how the different aspects of the subject were supposed to fit together to form the complete subject of General Relativity. This outcome is not surprising, as the participants would not be expected to have attained the higher levels of *cognitive process* outlined by Bloom's Revised Taxonomy after only initial study of the subject of General Relativity.

Discussion – Teaching General Relativity

Investigating the idea of giving feedback to the participants taking the GRCI was not an aim of this study at the outset. However, the participants discussed feedback at length during the interviews. All of the interviewed participants from the current study felt that it would be useful to receive feedback after answering the GRCI questions. This aligns with the idea that students feel as if feedback is useful in general, which is an idea that is well documented in the literature (Brown & Glover, 2006; Kluger & DeNisi, 1996; Zhu et al., 2020). Expanding on this, the participants were interested in finding out whether their answers to the GRCI questions were correct. In particular, the participants wanted to know where they went wrong, in order to highlight gaps and weaknesses in their understanding. Use of feedback in this manner is an example of self-regulated learning (Nicol, 2007), in which students reflect upon their performance and identify which topics they struggle with. This self-reflection can then provide a basis to guide the future study of the student. This is an example of how feedback can be used to encourage and support student-oriented learning.

In the context of the current study, it is worth recalling that concept inventories do not usually provide test-takers with feedback about their performance upon completion. However, the idea of combing feedback with concept inventories has previously been explored (Parker et al., 2022). General Relativity is an advanced and complex subject, and the free-response format of the GRCI questions provides flexibility for discussion between instructors and test-takers to take place. As such, feedback could be particularly useful for the GRCI. This potential direction was highlighted during the interviews with participants about their experiences of doing the GRCI. The participants noted that idea of getting feedback could open the way for new teaching interventions based upon the GRCI. Drawing on their own experiences, participants proposed that discussion of their answers could be the focus of tutorials with smaller groups or problem classes. This approach to using the GRCI could provide context for the teaching of General Relativity, which could facilitate students' learning of the subject.

The literature details several attempts made to teach General Relativity using other context-based approaches. Examples include the development of an interactive tool for investigating the geodesics of different spacetimes (Muller & Frauendiener, 2011); the use of workshop sessions to teach about curved spacetimes using black hole models (Zahn & Kraus, 2014); and the application of a teaching activity which made use of gravitational wave data collected from the LIGO installations (Burko, 2017). Within a conceptual context, Kaur et al. (2017) developed an approach based on analogies for introducing General Relativity concepts at the school level. The authors were motivated to take this approach because Einsteinian physics has a range of important real-world applications that students encounter in the everyday world. These examples indicate that contexts exist where the GRCI could also be used to assist with teaching. Further development of the GRCI in the teaching context could be an avenue for future research.

The findings from this study have important implications for the development of concept inventories for General Relativity and other mathematical physics topics. When authoring the GRCI questions, it was found to be difficult to disentangle the mathematics from the physics and concepts underpinning the theory of General Relativity. This would be expected, since there is a high level of interplay between these three aspects in the formulation of General Relativity. Similar issues were encountered during the development of concept inventories for the mathematical physics topics of Quantum Mechanics (Dick-Perez et al., 2016) and Electromagnetism (Baily et al., 2017; Ding et al., 2006), where the authors noted that testing for conceptual understanding of these subjects is challenging because of the amount of mathematics involved. This correlates with the findings from the current study in the context of Bloom's Taxonomy (Bloom, 1956; Krathwohl, 2002), where appreciation of the mathematical, physical and conceptual dimensions of General Relativity are required to demonstrate understanding and mastery of the subject.

LIMITATIONS

The process used to develop the GRCI questions was largely successful. However, potential limitations were also identified. First, the study involved a total of 26 participants. It would have been useful to have a greater number of participants answer the GRCI questions and be interviewed about their experiences. Having said this, the study was *proof of concept* in scope, with the aim of highlighting important points pertaining to the feasibility and development of the GRCI. The smaller number of participants was sufficient to complete this particular objective. Additional work to expand the potential scope of the study and further investigate the points highlighted by the study is encouraged.

Second, there were instances in the GRCI questions where students were scaffolded towards giving a particular answer. This included the use of key words and the instructions given in the questions. That said, General Relativity is a complex subject which requires a certain level of detail and context for questions to be posed about it. This means that some amount of scaffolding may be a necessary feature to properly set up the GRCI questions.

Finally, General Relativity is a broad subject, which invertedly gives a subjective dimension to the process of selecting key concepts to base the GRCI questions on. This issue arises because experts from different areas of General Relativity, such as theoretical physicists and observational cosmologists, might select different concepts to be the key concepts of the subject. This could potentially lead these experts to develop versions of the GRCI that contain very different questions, based upon their own research expertise and academic biases. Despite this, there should still be a number of concepts that all experts in the field of General Relativity agree upon to be fundamental to the formulation and understanding of the subject. This consensus-based idea was employed when selecting the concepts to base the questions on the GRCI upon.

CONCLUSIONS

The study investigated two research questions:

RQ1: How can a concept inventory containing free-response questions be used to test conceptual understanding of General Relativity?

RQ2: What thought processes are used by students to answer the GRCI questions, and how do students react to the conceptual nature of the questions posed in the GRCI?

The study was based on using free-response questions to develop a concept inventory for the mathematical physics subject of General Relativity. The aim of the study was to highlight important points related to the feasibility and development of the GRCI. The study involved a total of 26 participants, and was *proof of concept* in scope. Data were collected by having the 26 participants work through the GRCI, and by conducting interviews with four of the participants about their experience of answering the questions on the GRCI.

In order to answer **RQ1**, the GRCI questions were developed and tested. It was noted from the literature that concept inventories are difficult to develop for mathematical physics topics because of the interplay between mathematics, physics and concepts. As such, identifying the conceptual grounding of General Relativity taken as the fundamental starting point of the study. Concepts which formed the foundations of General Relativity were used to draft questions for the GRCI. Getting expert review was found to be an essential part of the process, as this reduced the potential for academic bias in the content of the questions. No issues were found when students worked through the GRCI. The students understood what the GRCI questions were asking, and provided answers to them in the intended way. Consequently, this limited pilot study provided encouraging indicators for the potential use of the GRCI in a teaching context. Moreover, these findings from the process used to develop and test the GRCI could facilitate future attempts to develop concept inventories for other mathematical physics topics.

RQ2 contained two parts. In order to answer the first part of **RQ2**, the written responses to the GRCI questions were analysed. This investigation revealed that students were less successful when answering questions based on the physical ideas of *Frames of Reference* and the *Principle of Equivalence*, while they were more successful when answering questions based on the mathematical ideas of *Curvature* and *Einstein's Field*

Equations. Further, the students tended to struggle to answer questions that required them to interpret results from General Relativity in a Cosmology setting. Overall, this demonstrated that the students were better at answering questions which required mathematical thought processes, but were less effective when answering questions which required conceptual thought processes and physical interpretations of the theory. These findings were consistent with the observation that the majority of General Relativity courses teach the subject using predominantly mathematical approaches. This highlighted the potential use of the GRCI in a teaching context, as it encourages students to consider the subject from a physical and conceptual standpoint, while retaining appreciation of the mathematical characteristics of General Relativity.

In order to answer the second part of **RQ2**, the data gathered from the GRCI interviews were analysed. The interviews revealed that the students felt that the questions on the GRCI were suitable for testing conceptual understanding of General Relativity. Further, the students contrasted the conceptual questions on the GRCI with the mathematical questions which they were familiar with from their previous studies of General Relativity. Students proceeded to discuss the nature of General Relativity as a difficult and abstract subject, and this gave rise to the idea of using General Relativity in science engagement ventures. Students reflected upon their individual performance on the GRCI questions, and felt that receiving some level of feedback from the GRCI would facilitate their learning and understanding of the subject. In summary, the interviews allowed students to demonstrate their knowledge, highlight their weaknesses and stimulate discussion about the subject of General Relativity. These findings indicate that the GRCI is a useful educational instrument which encourages students to consider the conceptual dimension of General Relativity, and it could be implemented to support the learning and understanding of General Relativity in a teaching context.

The findings from answering **RQ1** and **RQ2** highlight the possible use of the GRCI in a teaching context. Mathematical physics topics such as General Relativity contain physical, mathematical and conceptual dimensions, and grasping the interplay between these different characteristics is integral to gain an understanding of the subject. An instrument such as the GRCI helps to bridge this gap in understanding by having students consider the different aspects of General Relativity. Combining instruments similar to the GRCI with feedback could encourage the development of student-oriented teaching approaches. The process used to develop the GRCI could facilitate further efforts to develop concept inventories for mathematical physics topics. Looking ahead, such instruments could prove to be useful in the teaching context for other mathematical physics topics. Taken together, these findings are the main research output of this work.

Author contributions: MAJP: designed the research method, conducted the research, and wrote the article body text; **HH, NSJB, SEJ:** supervised the research and provided corrections and comments to improve the article body text. All authors approved the final version of the article.

Funding: This article was supported by the Open University.

Acknowledgements: The authors acknowledge the helpful contributions of the academics who reviewed the GRCI questions. This was an integral part of the development process, and was essential to the success of the research. Further, the authors owe a debt of gratitude to the participants who answered the questions on the GRCI, and to the participants who were interviewed about their experiences of doing the GRCI. Without the generous contribution of these participants, the research could not have been completed.

Ethics declaration: This study was approved by the Open University Human Research Ethics Committee on 31 July 2017 with approval code HREC/2017/2629/Parker/1. Students gave consent (via consent form) for their involvement of the study before taking the test. The consent form informed them that data would be used for research into the development of physics assessment. Data protection was implemented as follows. The author team were the only ones who had access to the response and interview data, which were held by the corresponding author on his personal computer. Interview data were gathered using a recording device, which belonged to the lead author. The data were carefully stored such that they could not be accessed by anybody outside the author team. The recording device was kept in a locked cabinet, and the personal computer with the response data was locked whenever the corresponding author was away from it.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analysed during this study are available from the PhD thesis "Establishing physics concept inventories using free-response questions", which can be found at: https://oro.open.ac.uk/73254/.

REFERENCES

Aslanides, J. S., & Savage, C. M. (2013). Relativity Concept Inventory: Development, analysis, and results. *Physical Review Special Topics Physics Education Research, 9*, Article 010118. https://doi.org/10.1103/PhysRevSTPER.9.010118

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

Baily, C., Ryan, Q. X., Astolfi, C., & Pollock, S. J. (2017). Conceptual assessment tool for advanced undergraduate electrodynamics. *Physical Review Physics Education Research, 13*, Article 020113. https://doi.org/10.1103/PhysRevPhysEducRes.13.020113

Bandyopadhyay, A., & Kumar, A. (2010). Probing students' understanding of some conceptual themes in general relativity. *Physical Review Special Topics Physics Education Research, 6*, Article 020104. https://doi.org/10.1103/PhysRevSTPER.6.020104

Bloom, B. (1956). *Taxonomy of educational objectives: The classification of educational goals*. McKay.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., Clarke, V., & Terry, G. (2014). Thematic analysis. *Qualitative Research in Clinical Health Psychology, 24*, 95–114.

Brown, E., & Glover, C. (2006). Evaluating written feedback. In C. Bryan, & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 81–91). Routledge.

Bulut, O., Cutumisu, M., Aquilina, A., & Singh, D. (2019). Effect of digital score reporting and feedback on students' learning in higher education. *Frontiers in Education, 4*. https://doi.org/10.3389/feduc.2019.00065

Burko, L. M. (2017). Gravitational wave detection in the introductory lab. *The Physics Teacher, 55*, 288–292. https://doi.org/10.1119/1.4981036

Catalyst Harvard. (2024). *Mixed methods research*. https://catalyst.harvard.edu/community-engagement/mmr/

Conlon, M., Coble, K., Bailey, J. M., & Cominsky, L. R. (2017). Investigating undergraduate students' ideas about the fate of the Universe. *Physical Review Physics Education Research, 13*, Article 020128. https://doi.org/10.1103/PhysRevPhysEducRes.13.020128

Dick-Perez, M., Luxford, C. J., Windus, T. L., & Holme, T. (2016). A quantum chemistry concept inventory for physical chemistry classes. *Journal of Chemical Education, 93*(4), 605–612. https://doi.org/10.1021/acs.jchemed.5b00781

Ding, L., Chaby, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research, 2*, Article 010105. https://doi.org/10.1103/PhysRevSTPER.2.010105

Hartle, J. B. (2008). General relativity in the undergraduate physics curriculum. *American Journal of Physics, 74*(1), 14–21. https://doi.org/10.1119/1.2110581

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141–158. https://doi.org/10.1119/1.2343497

Hufnagel, B. (2002). Development of the astronomy diagnostic test. *Astronomy Education Review, 1*(1), 47–51. https://doi.org/10.3847/AER2001004

Institute of Physics. (2024). *Degree accreditation and recognition*. https://www.iop.org/education/support-work-higher-education/degree-accreditation-recognition

Kaur, T., Blair, D., Moschilla, J., Stannard, W., & Zadnik, M. (2017). Teaching Einsteinian physics at schools: Part 1, models and analogies for relativity. *Physics Education, 52*(6), Article 065012. https://doi.org/10.1088/1361-6552/aa83e4

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical view, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

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

Muller, T., & Frauendiener, J. (2011). Studying null and time-like geodesics in the classroom. *European Journal of Physics, 32*, 747–759. https://doi.org/10.1088/0143-0807/32/3/011

Nicol, D. (2007). E-assessment by design: Using multiple-choice tests to good effect. *Journal of Further and higher Education, 31*(1), 53–64. https://doi.org/10.1080/03098770601167922

Parker, M., Hedgeland, H., Braithwaite, N., & Jordan, S. (2022). Student reaction to a modified force concept inventory: The impact of free-response questions and feedback. *European Journal of Science and Mathematics Education, 10*(3), 310–323. https://doi.org/10.30935/scimath/11882

Parker, M., Hedgeland, H., Braithwaite, N., & Jordan, S. (2023). Establishing a physics concept inventory using computer marked free-response questions. *European Journal of Science and Mathematics Education, 11*(2), 360–375. https://doi.org/10.30935/scimath/12680

Porter, L., Taylor, C., & Webb, K. (2014). Leveraging open source principles for flexible concept inventory development. In *Proceedings of the 2014 Conference on Innovation Technology in Computer Science Education* (pp. 243–248). https://doi.org/10.1145/2591708.2591722

Rebello, N., & Zollman, D. (2004). The effect of distractors on student performance on the force concept inventory. *American Journal of Physics, 72*, 116–125. https://doi.org/10.1119/1.1629091

Semon, M. D., Malin, S., & Wortel, S. (2009). Exploring the transition from special to general relativity. *American Journal of Physics, 77*, 434–438. https://doi.org/10.1119/1.3088883

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

Stannard, W., Kersting, M., Kraus, U., & Moschilla, J. (2017). Research into the teaching and learning of Einsteinian physics in international contexts. In *Proceedings of the 2017 GIREP-ICPE-EPEC Conference*.

Thornton, R., & Sokoloff, D. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics, 66*(4), 338–352. https://doi.org/10.1119/1.18863

University of Auckland. (2024). *Thematic analysis*. https://www.auckland.ac.nz/en/science/about-the-faculty/school-of-psychology/psychology-research/studies-methods-and-ethics.html

Zahn, C., & Kraus, U. (2014). A toolkit for teaching general relativity: I. Curved spaces and spacetimes. *European Journal of Physics, 35*, Article 055020. https://doi.org/10.1088/0143-0807/35/5/055020

Zeilik, M. (2003). Birth of the astronomy diagnostic test: Prototest evolution. *Astronomy Education Review, 1*(2), 46–52. https://doi.org/10.3847/AER2002005

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behaviour and learning gains in formative assessment of scientific argument writing. *Computers and Education, 143*, Article 103668. https://doi.org/10.1016/j.compedu.2019.103668

◆❖◆