GR in VR: Using Immersive Virtual Reality as a Learning Tool for General Relativity

Kristen Schumacher, University of Illinois, Urbana-Champaign

Kristen Schumacher is currently a graduate student in physics with a research focus on modified theories of gravity. She founded the POINT project (Physics Outreach and Instruction through New Technologies) at UIUC to bring these abstract concepts to a wider audience through virtual reality.

Sonali Joshi, University of Illinois, Urbana-Champaign

Sonali Joshi is a graduate student in physics with a research focus on effective models for condensed matter systems. She is the head of the POINT (Physics Outreach at Illinois through New Technologies) VR Production Team. The VR simulations developed by POINT focus on conveying general relativity concepts.

Jina Kang, University of Illinois, Urbana-Champaign

Dr. Jina Kang is an Assistant Professor in the Department of Curriculum and Instruction at the University of Illinois Urbana-Champaign. Her research centers on three interrelated areas: data science, data-driven design, and STEM education. She applies different analytical approaches of data science to understand students' scientific inquiry behaviors in technology-enhanced learning environments and investigates data science skills required in science disciplines.

Eric Shaffer, University of Illinois, Urbana-Champaign

Eric Shaffer is a Teaching Associate Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. He also serves as the Associate Director for Education at the Center for Immersive Computing at that same institution. He teaches a revolving set of courses including Game Development, Computer Graphics, and Scientific Visualization. His research interests include using visualization for education, data-centric scientific computing, and geometric modeling.

Jessica Raley, University of Illinois, Urbana-Champaign

Jessica Raley is the outreach coordinator for the Illinois Center for Advanced Studies of the Universe (ICASU) at the University of Illinois Urbana-Champaign. She is also the adviser for the P.O.I.N.T. VR program.

Jose Nijaid Arredondo, University of Illinois, Urbana-Champaign

Graduate student working on the gravitational wave modeling of inspiraling compact binaries, with an emphasis on the effects of eccentricity and spin precession, and interested in expanding the teaching of general relativity to broader audiences, including K-12 curricula.

Brandon Mark Buncher, University of Illinois, Urbana-Champaign

I am am a physics Ph.D. student studying cosmology with a focus on large scale structure, weak gravitational lensing, big data analysis, and machine learning.

Rajan Patkar, University of Illinois at Urbana-Champaign

Rajan Patkar is a sophomore in computer engineering at the University of Illinois Urbana-Champaign. His goal is to connect his core engineering skills with fields in which he can make cross-disciplinary connections and accelerate progress, like physics education. He joined the POINT team (Physics Outreach and Instruction through New Technologies) to focus on the outreach and research efforts.

Katherine Zine, University of Illinois, Urbana-Champaign

I am a first year physics graduate student at the University of Illinois at Urbana-Champaign. I did my undergraduate at Indiana University Bloomington. My research is in theoretical nuclear physics on neutron star equations of state.

Daniel Alfredo Caballero, University of Illinois, Urbana-Champaign

I am a colombian Physics PhD. student at UIUC doing research in classical gravity.

Dr. Alexandria Tucker, University of Illinois, Urbana-Champaign

Dr. Alexandria Tucker is an NSF MPS Ascend Postdoctoral Fellow at the University of Illinois in Urbana-Champaign.

Mireille Tan, University of Illinois, Urbana-Champaign

I am a freshman in B.S. in Computer Science, with current research interests in Human-Computer Interaction and Computer Graphics. For this paper, I implemented several features related to user interface and experience for our simulation in Unity.

Christopher Vistian, Physics Outreach at Illinois Through New Technologies

Student of Mathematics and Computer Science at the University of Illinois at Urbana-Champaign.

GR in VR: Using immersive virtual reality as a learning tool for general relativity

Abstract

According to general relativity, gravity can be understood as a curvature of spacetime in response to the presence of matter and energy. Students often struggle to visualize the geometry of curved spacetime. The standard demonstration used to aid in visualization, that of a ball on an elastic sheet, is fundamentally flawed and may lead to misconceptions. Recent research suggests that virtual reality can improve understanding of spatially complex or abstract concepts. We hypothesize that an interactive virtual reality demonstration involving masses in a curved 3D spatial grid, with clocks representing the relative passage of time, would support improved conceptual understanding and impact attitude among students learning general relativity compared to traditional methods. To test this hypothesis, undergraduate students with no formal experience with general relativity are recruited to evaluate the virtual reality simulation. The students first take a questionnaire to determine a baseline for their conceptual understanding of general relativity, with confidence-scaled multiple choice and written response questions. The experimental group experiences an interactive virtual reality demonstration in which the subjects can move objects through space and time to visualize how mass curves spacetime. An instructor leads the control group through the standard ball on a sheet demo while delivering content orally. Students in both groups are prompted by an instructor to explore the relationship between mass, gravity, and time, guided by a set of conceptual questions. Immediately after the demonstrations, students complete the same questionnaire and a survey about learner attitude and simulation usability. Few previous studies focus on the conceptual understanding of general relativity, and even fewer examine the possibility of immersive learning as a tool for teaching this topic. Our work addresses this gap by designing a novel immersive technique for visualizing relativistic effects and comparing this technique to existing non-immersive methods of instruction.

Introduction

Albert Einstein's theory of general relativity (GR) is currently the leading theory to explain gravity, one of the four fundamental forces of physics. First proposed in 1915, GR has since been validated by numerous experiments, including the detection of gravitational waves that captured the public's attention in 2015 [1], [2]. In addition to the significant contribution of GR to our understanding of the universe, it also finds practical application in our daily lives through the GPS that allows us to navigate with our phones. Unfortunately, GR provides substantial conceptual and mathematical challenges, particularly for beginner students.

According to GR, mass and energy simultaneously curve space and time, which are mathematically described as a four-dimensional surface called spacetime. The curvature of spacetime is so far outside the realm of our daily experience that visualizations are essential to understanding. One of the most common visualizations for teaching GR is the "ball and sheet" demonstration. A heavy ball is placed on a taut elastic sheet, causing it to sag. Students are taught that this represents a massive object curving the "fabric" of space and time. However, “most features of this model are some mixture of incorrect and misleading” [3]. Little previous research has been done on the conceptual understanding of general relativity, and even less has examined the possibility of immersive learning as a tool for teaching this topic. Our work addresses this gap by designing a novel immersive technique for visualizing relativistic effects and comparing this technique to existing non-immersive methods of instruction.

Virtual reality (VR) offers a promising new avenue for dynamic, three-dimensional, inquiry-based education—a method shown to be more effective than traditional lectures [4]. Since VR is unconstrained by physical limitations, it provides physics educators with a unique tool for helping students visualize and understand the most abstract physics concepts. One of the key contributions of our work is a new VR simulation to help students develop an understanding of GR. This simulation, developed by *Physics Outreach and Instruction through New Technologies* (POINT), covers some of the most important concepts of GR and is intended to replace the traditional ball and sheet demonstration in formal classroom settings or for general public engagement. Once finished, this simulation will be available for free on the Meta Quest store, and the code will be open-source through a public GitHub.

In addition to this software, we also contribute a set of survey questions that may be used to evaluate college-level students’ basic conceptual understanding of GR. There is currently no single standard assessment instrument in the physics education community for measuring understanding of GR concepts. We hope that the development and public release of our survey can aid others in developing similar assessments.

Finally, we provide a report on an initial pilot study comparing the effectiveness of our VR simulation with traditional methods for learning GR. For this purpose, 20 undergraduate students are recruited to participate in either VR or non-VR demonstrations. Each group is given a survey before and after their experience. This “pre-post” survey contains confidence-scaled multiple choice, free response, and Likert-scaled attitudinal questions. The survey results are used to compare the groups’ pre-post demonstration attitudes and understanding. From the conceptual multiple choice questions, we find a small improvement in understanding for the VR group over the control group. All participants across both groups report an increase in confidence after the experience, while the attitudinal changes from pre-post are not significant. The free response answers provide a better window into the participants’ thought processes. Overall, the results of this pilot project indicate that our method shows promise as a learning tool for GR.

Background

General Relativity and Teaching Methods

Most people conceptualize gravity based on their everyday interaction with it. Thus, gravity is commonly viewed as a force that causes objects to fall to the ground instead of floating into space. This limited interpretation of gravity was formalized by Isaac Newton, who described gravity as a force that attracts all massive objects to one another. However, Newtonian gravity doesn’t account for more extreme scenarios (e.g. spacetime around highly massive objects), where GR becomes applicable. According to GR, the trajectories of light and matter in curved spacetime create the effect of gravity commonly attributed to the force between objects in the classical Newtonian model. Considering how significantly GR’s four-dimensional geometrical representation of gravity differs from our lived experience, successfully teaching the theory to students without using complex mathematics is notoriously challenging.

For example, the concept of time dilation, the slowing of time in the presence of mass and energy, is far outside the realm of daily experience. This concept contradicts Newton's theory of an "absolute time" that is the same everywhere in the universe and progresses at a constant rate. As a result, time dilation is commonly misinterpreted as a change in one's *perception* of time. However, in reality, the rate at which time passes *fundamentally changes* in the presence of mass and energy. This concept can confuse first time learners of GR.

Unfortunately, the ball and sheet demonstration commonly used to explain GR has some drawbacks. Spacetime is represented as two-dimensional, even though there are actually three spatial dimensions and one temporal dimension. Therefore, the demonstration does not correctly depict how spacetime curves in response to massive objects. Additionally, this demonstration can lead to misconceptions about the direction in which objects fall in a gravitational field. Since the sheet is depressed due to Earth's gravitational field, users can only place objects on top of the sheet, limiting how many possible orientations of the objects can be explored. Furthermore, if students first place a heavy object and then a lighter object on the sheet, the lighter object will always roll "down" to the heavier one. However, in space, there is no preferred direction, and objects in a gravitational field fall to the center of mass. For a summary of the strengths and weaknesses of the ball and sheet demonstration, see [5].

The connection between gravity and time dilation is also unclear in the ball and sheet demonstration. Thus, if including a discussion of time dilation, the instructor must describe the effect verbally without a visual component to illustrate this connection. Reference [5] points out that "the time dimension tends to be a neglected feature" in this analogy. They also find that students struggle to understand that curvature and movement in spacetime equate to curvature and movement in *both* space and time or that objects are moving continuously along the time dimension.

To avoid these issues, a more complete visualization of spacetime curvature requires a three-dimensional spatial model that viewers can immerse themselves in, as well as a demonstration of time dilation. One method for providing this immersive experience is through VR. VR allows the user to view grid lines representing the deformation of distances in three dimensions, accurately showing how space responds to the presence of matter and energy. While VR is visually limited to three dimensions, time dilation can be illustrated in other ways. For example, in VR analog clocks and a ticking noise can demonstrate the passage of time to build a full picture of what happens to *spacetime*, not just space.

Existing Literature

Limited research has been done on the best way to teach the concepts of GR to students. Until very recently, most high school physics instruction covered only classical theories of gravity [5], [6]. When schools do introduce relativity, it is often through the theory of special relativity (SR), which describes how speed affects mass, space, and time in the absence of gravity. There have been efforts to evaluate methods for teaching SR to students [7]–[12]. However, as a special case of GR that does not include the effects of gravity, metrics that have been developed to measure student understanding of SR, like the Relativity Concept Inventory presented in [13], do not encapsulate all of the same concepts that one needs to measure to ascertain an understanding of GR.

Although discovered over a century ago, it is only in the last decade that educators began to introduce GR to middle and high school curricula and study student's comprehension of the topic [5], [14]–[18]. These recent efforts to introduce GR at a younger age suggest that students benefit from an earlier conceptual understanding of GR, as this may help them to grasp the complex concepts of spacetime later in life [19]. Students as young as 10 years old are capable of statistically significant improvement in conceptual understanding of GR after 6 in-class lessons [14]. However, GR is usually not introduced until the undergraduate or graduate level. Even for undergraduate understanding, studies are scarce [20]–[22].

The merits of the common ball and sheet demonstration are widely debated by physicists and science educators but rarely quantified systematically. Additionally, when it has been studied, this demonstration has been used to teach aspects of classical gravity rather than GR [23], [24]. A "systematic metaphor analysis of textbooks and research literature," first performed by [5], also analyzes high school students' application and understanding of the analogy to explain GR. As described in that paper, there are several ways that this demonstration can lead to misconceptions. The results of [5] are confirmed by [25], but [26] find that some of these misconceptions might be mitigated with a detailed discussion about the limitations of the ball and sheet demonstration.

One novel method that might be employed to communicate the concepts of GR is VR. Numerous studies have been conducted on the effectiveness of VR as a learning tool. A meta-analysis of 27 of these studies finds that virtual worlds had a positive impact when used with learning [27]. Another study finds VR interventions effective for students with varied learning styles [28]. Additional studies compare the effectiveness of VR interventions to demonstrations using previously existing technologies for various subject areas. For example, [29] compares the impact of immersive VR to that of a slideshow. They find the VR simulation increases engagement and excitement, but students learn more from the slideshow.

Preliminary work developed rendering methods for visualizing GR and other relativistic effects using large-screen immersion [30]. However, portable VR devices were not used, and implications for enhanced learning were not investigated.

Despite the growing body of literature on teaching GR and the more established research about using VR as a learning tool, few studies explore both. This dearth of literature is in part due to the lack of VR simulations that target these specific topics. To the authors' knowledge, only four simulations currently exist, including our own. One interactive VR simulation, *Spheres*, features three astrophysics modules, including one that allows viewers to fall into a black hole and observe the generation of gravitational waves [31]. Another simulation, *Black Holes: Light & Matter*, demonstrates gravitational lensing around a black hole and explores the different orbits that one can produce near a spinning black hole [32]. A third, *SciVR*, is a smartphone application with multiple interactive VR modules that explore topics in astronomy and astrophysics [33]. The most relevant module to our work shows the generation of gravitational waves and how these waves appear in a detector.

Of these simulations, only *SciVR* has been studied as a tool to communicate science concepts. This study did not cover gravitational topics with participants [34]. Furthermore, the focus of each of the above programs is to visualize a specific topic within GR rather than to teach fundamental concepts, such as the nature of spacetime and how its curvature relates to gravity.

Ball & Sheet Demonstration	VR Demonstration
Uses a 2D grid to represent a 4D spacetime.	Uses a 3D grid and clocks to represent a 4D spacetime.
Unclear that the curvature of spacetime is the same on every side of the object.	The curvature of spacetime is the same on every side of the object.
No visual representation of time dilation.	Time dilation represented with clocks.
Objects only fall "down."	Objects can be released from any orientation and always fall to the center of mass.

Table I: A list of the drawbacks of the ball and sheet demonstration and how the VR simulation addresses them.

With this pilot project, we introduce a new simulation targeted at these fundamental concepts and explore the existing gap in the literature.

Study Design and Research Questions

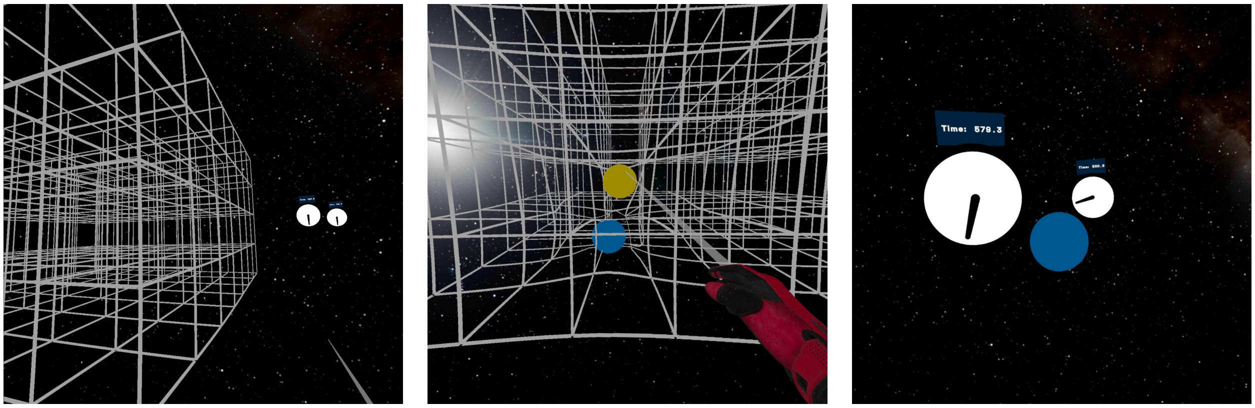
The purpose of this study is to investigate the benefits and disadvantages of using VR to communicate abstract concepts in GR that may be hard to visualize. To do this we compare two different demonstrations: a traditional physics demonstration and a new VR simulation. Study participants for both demonstrations fill out a pre-post survey designed to address specific research questions.

Study Design

Ball and Sheet Demonstration: The traditional demonstration used to illustrate gravity as a curvature of spacetime features a two-dimensional elastic sheet stretched over a frame. Participants are guided through the demonstration by an instructor with a script. The sheet represents spacetime, and an object placed on the sheet, such as a heavy ball, represents an astrophysical object (e.g., a star). The weight of the object causes the sheet to dip, providing a visible curvature that is used as a model for the curvature of spacetime in response to matter. One benefit of this demonstration is that the amount of curvature is directly related to the mass of the object placed on the sheet, which is a good analogy for what happens in GR. Participants can also see how this curvature affects other objects placed on the sheet. The first object warps spacetime, creating a gravitational effect on other objects, which fall into the center.

Virtual Reality Simulation: POINT has developed a new VR simulation to communicate some of the fundamental concepts of GR and address the drawbacks of the ball and sheet demonstration (Table I). Participants are guided through this simulation using a similar script as in the ball and sheet demonstration. The simulation begins with the user standing in space, with a three-dimensional grid and two clocks visible (Fig. 1a). The hands of the clocks are rotating at the same constant rate, and a slight audible ticking sound represents the passage of time on the clock closest to the user. Together, the shape of the grid and the ticking of the clocks represent spacetime.

In front of the user is a yellow sphere, which the instructor describes as an object with the same mass as our sun. The user can pick up this sphere and move it into the three-dimensional grid, which curves in response to the location of the sphere (Fig. 2a). The amount of curvature of this grid is determined by the mathematical equations of GR (see Appendix A). Hence, the curvature at a particular point in the grid depends on the mass of the object and the distance from the object. This visualization is more physically accurate than the traditional analogy of a



(a) The grid on the left is a representation of space and the clocks on the right show the passage of time.

(b) Each massive object curves the grid. When released they move towards each other along the curvature.

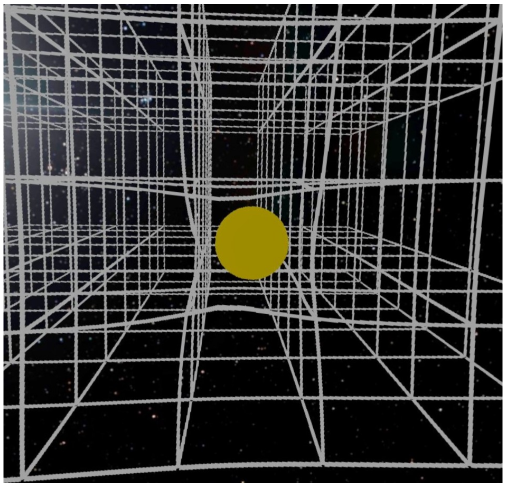
(c) The clocks slow down in the presence of massive objects to demonstrate time dilation.

Figure 1: Inside the VR simulation.

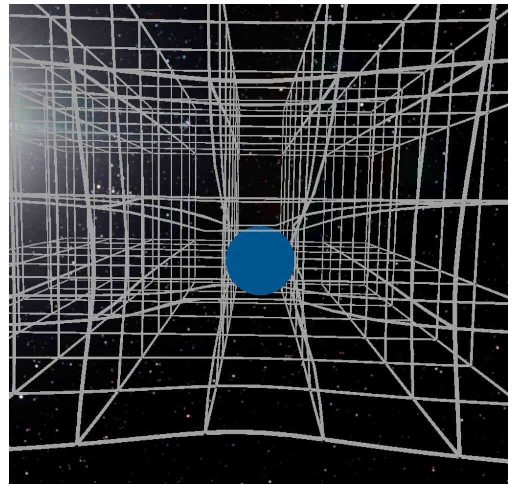
ball on a sheet. Additionally, bringing the object close to one of the clocks will cause that clock to slow down relative to the other (Fig. 1c). The rate of change is again determined by the mathematical equations of GR (Appendix A).

Once these concepts have been explored, the user can progress through the simulation by clicking on one of the menu items to select a different mass. The "Big Mass" option replaces the yellow sphere with a blue one, which the instructor describes as an object with a mass several times the mass of our sun. The designers of the simulation chose an object that was two times more massive than the previous object to make the difference very apparent. Moving this sphere in the three-dimensional grid makes it curve much more dramatically (Fig. 2b). Similarly, moving this sphere close to one of the clocks creates a greater difference in rates than the previous sphere did. The use of two different masses helps participants explore the connection between the amount of mass and the strength of gravity.

Finally, participants can select the "Two masses" menu item to work with both the yellow mass and the blue mass at the same time. Participants are encouraged to grab both objects, pull them apart, and simultaneously release them. The objects move towards each other, obeying the laws of gravity. Although the yellow object travels farther, both objects move. Hence, this simulation addresses the misconception that only the less massive object falls towards the more massive one, like an apple falling toward Earth. Another misconception is that the apple is falling "down." In reality, the apple is moving toward the center of the more massive object—the Earth. In our simulation, objects can be released from any orientation and will move toward each other, demonstrating that there is no preferred "downward" direction in space. Additionally, the user can throw one object to make it orbit the other. As long as the user does not throw the object too hard or too far away from the other object, it will be pulled in and naturally enter a physically accurate orbit. If the objects are within the three-dimensional grid, the user can see the orbit following the natural curvature of the grid.



(a) A smaller mass.



(b) A larger mass.

Figure 2: The massive objects curve the spacetime grid; the more massive the object, the more the grid curves. The curvature of the grid, when shown in a two-dimensional photo, looks slightly different than in the VR headset due to its curved screen.

Research Questions

A pre-post survey is given to all participants before and after the demonstrations. The questions included in the pre-post survey assess the following research questions:

1) Comparing participants using the traditional demonstration with those using the VR simulation, is there a measurable difference in comprehension on the following topics:
   a) the shape of spacetime curvature
   b) the relationship between time and gravity
   c) the direction objects move in a curved spacetime
2) Is there a difference in attitude or engagement between the two groups?
3) Does experiencing the immersive VR simulation improve participants' confidence in their answers relative to traditional methods?

The targeted concepts represent some of the main features of GR. These features are inherently different from everyday experience and are typically not taught in introductory physics courses. Moreover, the traditional demonstration can be misleading [5], [25]. Thus, we expect a greater benefit from using VR in these areas. Additionally, measuring engagement is important because better engagement of students tends to lead to better outcomes, both in conceptual understanding and in long-term involvement with the material [27]. Finally, by assessing confidence, we hope to separate good guesses from true comprehension of GR.

We hypothesize that an interactive VR demonstration involving masses in a curved three-dimensional spatial grid, with clocks representing the relative passage of time, would support improved conceptual understanding and positively impact attitude among students learning general relativity compared to traditional methods.

Methods

To test our hypothesis, we base our methodology primarily on [35] and [36]. We randomly split half of our undergraduate participants into a control group, which perform a traditional physics demonstration, and the other half into an experimental group, which engage with a VR simulation. We record their prior knowledge and attitudes with a pre-survey and compare these to a post-survey to measure learning gains and attitudinal changes. In the remainder of this section, we report on the participants' demographic information, followed by the materials and procedures of our study.

Participants

Participants are recruited from volunteer undergraduate students at the authors' institution, all of them majoring in a STEM discipline. Of the 20 total participants, 15 are physics majors, 2 are computer science majors, 2 are majoring in both mathematics and computer science, and 1 is an electrical engineering major. Half of these are freshmen, 3 are sophomores, 5 are juniors, and 2 are seniors. We require that participants have not had formal coursework in GR, but expect the majority to be acquainted with basic physics concepts, such as Newtonian gravity. Half of the participants are randomly placed in the control group, while the other half are placed in the VR group.

Of the $n = 20$ participants, 75% are male ($n = 15$), while only 15% are female ($n = 3$) and 10% are non-binary ($n = 2$). With such a small sample size, we defer gender-related analysis to future studies. Indeed, the recent study [37] finds slight gender differences in responses to conceptual questions while finding larger differences in responses to attitude questions.

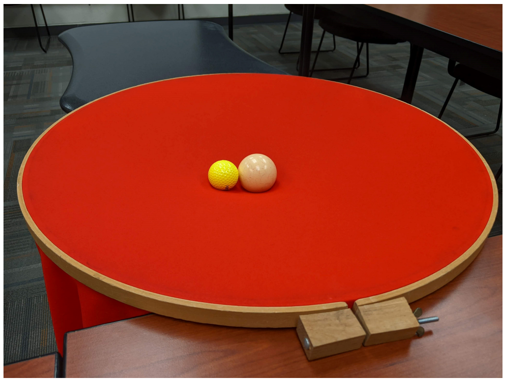
As part of the pre-survey, participants are asked about their prior experience with VR and GR. 60% of participants report they have used some type of VR headset before, although only 10% of participants report their VR experience is "extensive." Although participants have had no formal training in GR, they report varying degrees of informal experience with GR; 80% of participants report "little" or "no" experience with GR and only 20% report "some" experience.

Materials

Our study places participants in one of two activities: a traditional ball and sheet demonstration or a new VR experience. The ball and sheet demonstration features a flexible cloth sheet stretched over a fabric hoop that is two feet in diameter. The fabric hoop is balanced between two flat supports so that most of the sheet is suspended and can dip under the weight of objects placed on it. Participants place a golf ball and then a billiard ball on the sheet, which are described by the experimenter as astrophysical objects with different masses. In the VR experience, spacetime is represented by clocks that slow down and a three-dimensional grid that curves near virtual spheres. This simulation runs on Meta Quest 2 headsets, and participants can interact with the simulation via handheld controllers. Images of both demonstrations can be found in Figures 1 and 3.

Procedures

Participants are randomly assigned to either the control group or the VR group before arrival at the study location. The study proceeds similarly for both groups. Each participant arrives for



(a) Traditional demonstration of different masses on an elastic fabric.

(b) The Meta Quest 2 VR headset and controllers.

Figure 3: Demonstration materials.

their individual time slot and begins with the pre-survey, where randomized numerical identifiers are used to match surveys. Following a script designed for this study, we give a brief verbal description of spacetime and the historical reasons for replacing Newtonian gravity with GR.

The participants then engage in their assigned activity, the ball and sheet demonstration or the VR experience. Both feature guided exploration prompted by researcher questions. For example, “What characteristics do you notice about the curvature of the grid?” or “Is the curvature the same from all directions?” Participants are asked to make predictions and to test those predictions.

Most of the language is consistent between the two scripts, but some necessary adjustments are made for the VR simulation. For instance, participants in the VR group are shown how to hold the VR controllers and are told what buttons to push to interact. Furthermore, given that there is no visual analogy for time dilation in the ball and sheet demonstration, researchers explain this effect verbally. In the VR group, participants are prompted to “bring the mass close to the clock” and asked, “Does proximity to the mass affect the passing of time, and if so, how?” In the VR script, no explicit statement is made that time slows down in a gravitational field; rather, students discover this concept on their own.

After finishing their respective activity, participants are given the post-survey. Upon completion, the surveys are collected and processed electronically to be analyzed. Once finished with the study, participants in the control group are given the option to see the VR demonstration.

Measures

Surveys are used to measure participants’ understanding, confidence, and attitude towards the subject. The pre-survey begins by asking some basic biographical questions including the participants’ year in college, major, and gender identity, which are used to identify trends in our data. Participants are then asked to rate their previous experience levels with both VR and GR on a five-point scale. Since neither group of participants has had formal experience with GR, their levels of prior knowledge and attitudes are expected to be similar.

Main Concept	Acceptable Responses
The universe consists of a 4D spacetime.	Use the word “spacetime,” indicate that space and time are connected, describe dimensions of spacetime.
Spacetime curves.	State that spacetime curves or any synonym. Also accepted “space curves.”
This curvature is gravity.	Say gravity is equivalent to, caused by, or causing the curvature of spacetime.
Gravity is related to mass.	State that mass and gravity are related.
Time slows down near massive objects.	State that time is affected by gravity. This could include “time dilation,” “time slows down,” or “time warps.”

Table II: A rubric to grade the free response question, “Explain how space, time, and gravity are connected.” One point is awarded for each of the main concepts addressed in the answer.

The survey includes multiple choice questions addressing the shape of spacetime curvature, the relationship between time and gravity, and the direction objects move in curved spacetime. These questions were developed specifically for this study, as the *Relativity Concept Inventory* only contains questions about SR and not GR [13]. For a complete list of survey items used, see Appendix B. The pre-post surveys can be compared between the two groups to see how the different demonstrations affected participants’ understanding of these topics.

Participants’ rate their agreement with nine statements on a standard five-point Likert scale to record their attitudes toward GR, physics, and science in general. These statements are largely drawn from [25] but modified for undergraduates and to reflect the focus on GR.

Additionally, following the example of [13], after each multiple choice question, participants are asked to rate their confidence in their answer on a five-point scale: “Guessing”, “Hesitant”, “Neutral”, “Confident” or, “Certain.” This metric may help differentiate between similar pre-post survey responses, distinguishing true understanding from lucky guesses. Confidence levels across the two groups of participants can be compared to see if one type of demonstration leads to more confident responses than the other.

To help assess the practicality of using the VR program instead of the traditional method, participants are asked a series of questions about the activities’ usability. These include rating how easy the equipment was to use and how interesting the demonstration was. These questions should help identify common issues with the VR simulation and determine if these can be improved in future iterations. Participants in both groups are also asked to “Leave any additional comments you have about the usability of the activity.”

Finally, the post-survey contains one open-ended free response question. Participants are asked to “explain how space, time, and gravity are connected” in their own words. The research team developed a five-point rubric to code these responses (Table II).

To analyze the survey responses, we utilize non-parametric tests to estimate the significance of differences between groups. Since the data is ordinal, it fails the presumptions of parametric tests like the t-test [38]. First, a Kruskal-Wallis H-test checks if the four batches of data (separated by pre/post and VR/control) are centered equally [39]. If the Kruskal-Wallis H-test rejects the null hypothesis that all the batches come from the same population, then a Mann-Whitney U-test determines if the unpaired groups (VR and control) are distributed similarly [40]. For the four batches of data described above, the Kruskal-Wallis H-test fails to reject the null hypothesis, so a Mann-Whitney U-test cannot be run, as it would overestimate the likelihood of a significant result [40]. Therefore, we run a Wilcoxon signed-rank test to examine if the paired sets (pre-post survey) are distributed differently [39], [41].

Non-parametric tests are underpowered and underestimate the likelihood of a significant result [42]. Given this fact and the small sample size, the critical p-value of 0.05 is more applicable for this analysis. For all tests, p-values less than 0.05 are taken to be significant [42]. To quantify the effect size of our study we use Cohen's d, which normalizes the difference in pre-post survey results by their pooled standard deviation [43]. Generally, Cohen's d indicates a study's impact–the larger the d-value of a survey question, the larger the effect of the study on participants. The effect size is considered small for $d \approx 0.2$, medium for $d \approx 0.5$, and large for $d \approx 0.8$. Note that d can be negative for a given question in the event that more students answered the question correctly in the pre-survey than in the post-survey. For a given question Q, we define Cohen's d [43], including Hedges' small sample size correction [44], as

$$d_Q = \frac{\mu_{\text{post}} - \mu_{\text{pre}}}{\sigma_{\text{pool}}} \left[\frac{\Gamma\left(\frac{N-2}{2}\right)}{\sqrt{\frac{N-2}{2}}\Gamma\left(\frac{N-1}{2}\right)} \right], \tag{1}$$

$$\sigma^2_{\text{pool}} = \frac{(n_{\text{pre}} - 1)s^2_{\text{pre}} + (n_{\text{post}} - 1)s^2_{\text{post}}}{N - 2},$$

where μ_{pre} and μ_{post} are the mean group scores for question Q, n_{pre} and n_{post} are the sample sizes of each group, and s_{pre} and s_{post} are the sample standard deviation of each group for the pre- and post- surveys respectively. Additionally, $N = n_{\text{pre}} + n_{\text{post}}$ is the pooled number of samples and σ^2_{pool} is the pooled sample variance. Equation (1) is applied to each group (VR and control) separately. Because N is small (in our case $N = 20$), we modify the traditional formula for d with Hedges' correction factor (in square brackets) [44].

Another metric for student learning is the Hake gain, which normalizes the changes in students' pre-post answers by their maximum possible improvement [45]. We calculated the Hake gain g for each question Q within each group using the equation

$$g_Q = \frac{\mu_{\text{post}} - \mu_{\text{pre}}}{1 - \mu_{\text{pre}}}. \tag{2}$$

However, according to [46], the Hake gain introduces a bias that renders it uninformative for researchers interested in the impact of interventions on student learning. Furthermore, they recommend the use of Cohen's d to avoid the bias introduced by Hake, since Cohen's d mitigates the bias towards higher pretest means. In agreement with the findings of [46], our g_Q values illustrate only the ceiling effect of our study, not the amount of student improvement. Thus, g_Q will not be used to evaluate our results. The results of the other tests and metrics are discussed in the following sections.

Results

We analyzed the survey to test our hypothesis: *an immersive VR simulation improves comprehension and attitude towards GR as compared to traditional methods.* To address comprehension, we performed a one-sided Wilcoxon signed test on the answers to our conceptual physics questions, which are listed in Appendix B. We used this test to quantify the amount of *improvement* in the correctness of survey responses (Table III).

For the control group, we found no significant improvements in the scores, and for the VR group, only two questions showed significant improvement. Q1 asks *"Is the curvature of*

Question	Control			VR		
	% correct in pre-survey	p-value	Cohen's d	% correct in pre-survey	p-value	Cohen's d
Q1	70%	0.290	0.202	30%	0.037	0.759
Q2	100%	N/A	N/A	90%	0.290	0.410
Q3	80%	N/A	N/A	90%	N/A	N/A
Q4	100%	N/A	N/A	100%	N/A	N/A
Q5	80%	0.683	-0.202	90%	0.290	0.410
Q6	80%	0.683	-0.202	100%	N/A	N/A
Q7	80%	N/A	N/A	60%	0.037	1.005
Average	84%			80%		

Table III: Table of results of the one-sided Wilcoxon signed test for the conceptual physics questions; a p-value/d-value of "N/A" indicates that the pre-survey responses did not differ from the post-survey responses for all subjects. Significant p-values ($p < 0.05$) are shaded. No questions from the control group showed a significant improvement, while there was significant improvement for two questions in the VR group. Notably, these questions were less affected by the ceiling effect. See Appendix B for the list of questions and answers.

spacetime the same on all sides of a spherical object?", while Q7 features four sets of images of spacetime curvature around two different masses and asks participants to choose the most accurate illustration. These are the two questions that best highlight the differences between GR and Newtonian gravity. To answer the question correctly, the participant must recognize that the image showing curvature from every side is the correct choice. Therefore, those in the control group have to understand that the curvature of the sheet in the ball and sheet demonstration is not accurate for a four-dimensional spacetime.

For the two statistically significant questions (Q1 and Q7) in the VR group, the results of Cohen's d-test (Table III) showed a large improvement in conceptual understanding. Due to the large values of Cohen's d for Q1 and Q7, we expect that the pre-post changes for these questions will remain significant with a larger sample size. For Q5 and Q6 in the control group, the d-value is negative because some respondents answered these questions correctly in the pre-survey and incorrectly in the post-survey.

The pre-survey scores in the control group were higher than the pre-survey scores in the VR group. Thus, while the VR group showed more improvement, we cannot be sure if this difference was because of a ceiling effect in the control group. Although the difference between the two demonstrations may not be as great as indicated in this pilot project, the VR simulation did help correct misconceptions in the VR group.

Overall, most of the questions had already been answered correctly by participants in both groups. Considering participants' STEM backgrounds, the high pre-survey scores are not surprising, even with minimal experience in GR. It is possible that younger participants, such as those in middle and high school who have more limited experience with the material, would show a more dramatic improvement. Alternatively, a larger sample size of college students, particularly those from outside STEM majors, may allow additional analysis by providing more statistical power and mitigating the ceiling effect.

The results may have also been affected by unclear question wording. Some participants expressed confusion about Q5—*"When you release a smaller mass in the presence of a larger mass, which direction does it fall?"*—because the relative positions of the masses were not

Question	Control			VR		
	Pre-survey mean	Post-survey mean	p-value	Pre-survey mean	Post-survey mean	p-value
Q1	2.4	4.4	0.003	2.5	3.7	0.004
Q2	3.3	4.9	0.003	3.3	4.7	0.001
Q3	2.3	4.4	0.006	2.7	4.9	0.038
Q4	4	4.9	0.019	4.3	4.9	0.003
Q5	3.9	4.6	0.028	3.9	4.9	0.003
Q6	4	4.1	0.398	4.2	4.9	0.01
Q7	3.6	4.6	0.006	2.9	4.8	0.001
Average	3.4	4.6		3.4	4.7	

Table IV: Table of one-sided Wilcoxon signed test results for the Likert scale confidence scores regarding the conceptual physics questions. Significant p-values ($p < 0.05$) are shaded. All questions showed a significant improvement in confidence scores aside from Q6 for the control group. See Appendix B for the list of questions and answers.

specified. Participants who interpreted the small mass as being above the larger one may have incorrectly answered, *"Both"* (i.e., down and towards the center of the larger mass). While this answer is true for some orientations, it does not hold in all cases and was graded as incorrect. For Q6—*"Imagine two objects with different masses being held apart in space. When they are released, what happens?"*—some participants may have chosen the correct answer choice, *"They both move,"* despite a misunderstanding about the direction of that motion. Several participants in the control group reported during the activity that the larger mass moves *away* from the smaller mass because of the recoil as the two balls collided. They did not seem to understand that this was just an artifact of the analogy, not a gravitational effect. Furthermore, some students in the control group answered Q5 and Q6 correctly in the pre-survey and incorrectly in the post-survey, as demonstrated by the negative Cohen's d-value. This decrease in scores occurred only in the control group, which suggests participants may have been misled by the ball and sheet demo. However, the insignificant p-values for Q5 and Q6 indicate that more data is required to form conclusions.

Participants' confidence ratings were also analyzed with the Wilcoxon signed test (Table IV). Although the majority of the questions were answered correctly in the pre-survey, the confidence levels averaged around "Neutral" for both control and VR groups. Both groups became significantly more confident after their respective demonstrations, even if their responses to the conceptual questions had not changed. However, this increase in confidence did not always come with an improvement in understanding. Confidence increased for all participants regardless of whether they were correct in the post-survey.

To further explore the differences in comprehension between groups, we used a rubric (Table II) to score the free response question from the post-survey: *"Explain how space, time, and gravity are connected."* Average scores for these responses (Table V) were very similar between groups, although the VR group listed slightly more of the main concepts from the rubric. Almost all participants mentioned that spacetime curves and that this curvature is gravity. They were less likely to describe the relationship between gravity and time or to name mass as the source of the spacetime curvature (note that mass was not explicitly asked about). Although the control group was told verbally about the relationship between gravity and time, while the VR group was not, 20% more of the VR participants discussed this relationship in their free response. While this result is promising, we need a larger sample size to determine if

Group	Description of spacetime	Spacetime curves	Curvature = gravity	Gravity-mass relationship	Gravity-time relationship	Total score
Control	0.6	1	0.9	0.4	0.5	3.4
VR	0.6	0.9	0.9	0.5	0.7	3.6

Table V: Average free response scores; points were assigned based on the rubric (Table II).

Statement	**Control**		**VR**	
	Pre-survey mean	Post-survey mean	Pre-survey mean	Post-survey mean
I think physics is interesting.	4.8	4.8	5	4.9
I enjoy learning new scientific concepts.	4.8	4.9	4.9	4.9
I prefer to learn physics through activities rather than through lectures.	3.9	3.9	3.5	4
General relativity is an important idea for modern science.	4.5	4.7	4.6	4.7
I want to learn more about general relativity.	4.7	4.8	5	4.9
I can understand new physics topics quickly.	3.9	4	3.8	3.9
I could be a good physicist.	3.9	4.1	3.9	4.2
Science is only for smart people.	1.7	1.7	1.9	2
Physics is useful and relevant to me.	4.5	4.6	4.6	4.8

Table VI: Table of average attitudinal scores. No further statistical inferences were made because a Kruskal-Wallis test of the four groups failed to reject the null hypothesis.

exploring time dilation in VR improves student comprehension.

Examining the language of the responses themselves gives a better picture of the participants' understanding. For example, one of the participants in the VR group who improved on both Q1 and Q7 gave the following well-written free response: *"The three dimensions of space and the one dimension of time provide a location for any object. Put together, space and time create a four-dimensional grid that every particle lies on. Gravity is the bending of this spacetime, distorting space and time around massive objects."* In contrast, the free response answer sometimes revealed gaps in understanding. One participant in the VR group answered more conceptual questions correctly in the post-survey, but reported *"A strong gravitational force appears to pull space towards it and to slow down the time around it. I am not too sure how space or time impact gravity or each other."* The phrasing indicates uncertainty and confusion, attitudes that were reflected in this participant's confidence scores. Furthermore, this participant indicated later that they struggled to adjust to the controls in the VR headset, so perhaps this impacted their experience and confidence.

We tested for significant differences in participants' change in attitudes (i.e., comparing averages from pre-post surveys) between the VR group and control group using the Wilcoxon signed test, Kruskal-Wallis H-Test, and Mann-Whitney U test. We found no significant difference between these groups. Comparing averages, the largest change was found for the VR group whose preference for learning physics "through activities rather than through lecture" increased towards "Agree" (Table VI). Yet both groups arrived at very similar attitudes across all questions, as may be expected for a group composed primarily of physics majors.

Participants were also given the opportunity to leave comments about the usability of the activity. Though we expected these responses to be most useful for improving future iterations of the VR simulation and did receive feedback about the controls, they yielded some other notable results. Two of the participants in the control group mentioned that they struggled to conceptualize time dilation with the ball and sheet demonstration. One reported, *"It was hard to see where time is involved with the grid. Although I understood how space distorts with gravity, it was hard to put together how time is impacted,"* and the other said, *"i [sic] enjoyed the simplicity...but I feel like I didn't really understand how time affected it? how [sic] do you add an extra dimension?"* Meanwhile, two participants in the VR group reported high engagement with the simulation. The first said, *"I really enjoyed it and would love to see similar activities in formal physics classes,"* while the second wrote, *"If I could access that on my home device I would spend a lot of time in there. It was genuinely extremely interesting to engage with."*

Conclusion

General relativity (GR) is a difficult subject for students to learn due to its conceptual and mathematical difficulty. Visualizing spacetime curvature poses a particular challenge. Interactive activities have been demonstrated to improve conceptual understanding; however, current methods, such as the ball and sheet demonstration, are oversimplified and may cause confusion. We present an immersive virtual reality (VR) simulation, which demonstrates spacetime curvature, emphasizing time dilation, to investigate its effectiveness as a learning tool for undergraduate students.

Improvement in understanding was measured via paired questions in a pre-post survey. Participants were also asked to report their confidence in their answers on a Likert scale. We found that participants in the ball and sheet demonstration (the control group) did not show a significant improvement, though their confidence scores increased for all but one question. Meanwhile, subjects in the VR group significantly improved on two of the questions with a large effect size and reported increased confidence for all questions. However, it is not clear whether improvements in confidence correlate to improvements in conceptual understanding. We also measured via Likert scale questions whether participants' attitudes toward physics changed, since increased interest in, or appreciation for, physics is beneficial regardless of increases in conceptual understanding. However, we were unable to perform statistical measures of significance for attitudinal scores.

The results of this pilot project suggest the need for improvement of the survey items. As there is no concept inventory for GR and testing materials are limited, the questions developed for this study address a gap in the literature. However, this first test of the survey items revealed some issues with question or answer choice wording that could be improved upon in future iterations. The survey items could be further strengthened through survey validation techniques. Additionally, all of our survey items featured conceptual questions and none targeted numerical problem solving. Future studies could address the question of whether students find numerical problem solving easier after using VR to understand the concepts. This could be done in collaboration with an introductory GR graduate course.

Although our study was conducted at a large R1 university with ample recruitment potential, our recruitment process encountered challenges that restricted the number and diversity of participants, limiting the generalizability and utility of our findings. One of the primary challenges was the time constraint on our recruitment process, which resulted in a small sample size of only 20 participants, 10 in each group. Additionally, limited advertising opportunities in courses offered by other departments led to a recruitment pool that was primarily drawn from undergraduate physics courses. As a result, we observed a ceiling effect in both the conceptual and attitudinal questions. Moreover, the underrepresentation of women in physics may have influenced the gender ratio of participants, as women constituted less than 25% of physics bachelor's degree recipients in 2020 according to the American Institute of Physics [47]. In the future, we aim to enhance our study by broadening and diversifying our participant pool. With sufficient time and improved advertising strategies, we are confident that this goal can be attained.

The development of the VR simulation used in this study provides educators and students with an immersive tool to experience advanced physics concepts. Once finished, the developers plan to make this simulation open-source so that any institution or individual with Meta Quest 2 headsets will have access. The Meta Quest 2 was chosen for being commercially available, relatively affordable, and highly interactive via handheld controllers. Some institutions already have VR labs that lend out these headsets. In fact, VR simulations for science teaching at an undergraduate level have already been implemented internationally in multiple schools [48]–[52]. The present simulation can be used in classes related to GR such as Modern Physics, graduate General Relativity, and possibly undergraduate Classical Mechanics. Moreover, the wide accessibility of the simulation is well-suited for outreach projects and middle or high school classrooms.

Another interesting avenue for future work is improving the simulation to better communicate GR concepts. We plan to explore alternate visualization techniques, such as using glyphs instead of a grid to represent the curvature of spacetime and drawing an object's trajectory in space to better illustrate the effect of gravity on its motion. We also plan to refine the user interface, adding guidance and simplifying the user experience to better support active learning within the VR environment.

Given a more diverse recruitment pool and potential improvements to the VR simulation, continued research could have a significant impact on how students learn GR. The positive impact already seen in our small study encourages us to continue using the interactive learning that VR offers to make abstract concepts in physics more accessible to young students.

Acknowledgements

K.S. would like to acknowledge that this material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE – 1746047. A.T. would like to acknowledge that this material is based upon work supported by the National Science Foundation Postdoctoral Research Fellowship under Grant No. 2212983.

References

[1] C. M. Will, *Theory and Experiment in Gravitational Physics*. Cambridge University Press, 9 2018.

[2] B. P. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett.*, vol. 116, p. 061102, Feb 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.061102

[3] R. H. Price, "Spatial curvature, spacetime curvature, and gravity," *American Journal of Physics*, vol. 84, no. 8, pp. 588–592, Aug. 2016. [Online]. Available: https://doi.org/10.1119/1.4955154

[4] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, “Active learning increases student performance in science, engineering, and mathematics,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8410–8415, May 2014. [Online]. Available: https://doi.org/10.1073/pnas.1319030111

[5] M. Kersting and R. Steier, “Understanding curved spacetime,” *Science & Education*, vol. 27, no. 7-8, pp. 593–623, Sep. 2018. [Online]. Available: https://doi.org/10.1007/s11191-018-9997-4

[6] E. K. Henriksen, B. Bungum, C. Angell, C. W. Tellefsen, T. Frågåt, and M. V. Bøe, “Relativity, quantum physics and philosophy in the upper secondary curriculum: challenges, opportunities and proposed approaches,” *Physics Education*, vol. 49, no. 6, pp. 678–684, Oct. 2014. [Online]. Available: https://doi.org/10.1088/0031-9120/49/6/678

[7] O. Levrini and A. A. diSessa, “How students learn from multiple contexts and definitions: Proper time as a coordination class,” *Physical Review Special Topics - Physics Education Research*, vol. 4, no. 1, Apr. 2008. [Online]. Available: https://doi.org/10.1103/physrevstper.4.010107

[8] K. Dimitriadi and K. Halkia, “Secondary students’ understanding of basic ideas of special relativity,” *International Journal of Science Education*, vol. 34, no. 16, pp. 2565–2582, Nov. 2012. [Online]. Available: https://doi.org/10.1080/09500693.2012.705048

[9] A. Velentzas and K. Halkia, “The use of thought experiments in teaching physics to upper secondary-level students: Two examples from the theory of relativity,” *International Journal of Science Education*, vol. 35, no. 18, pp. 3026–3049, Dec. 2013. [Online]. Available: https://doi.org/10.1080/09500693.2012.682182

[10] F. Kamphorst, M. Vollebregt, E. Savelsbergh, and W. van Joolingen, “Students’ preinstructional reasoning with the speed of light in relativistic situations,” *Physical Review Physics Education Research*, vol. 15, no. 2, Sep. 2019. [Online]. Available: https://doi.org/10.1103/physrevphyseducres.15.020123

[11] T. Hughes and M. Kersting, “The invisibility of time dilation,” *Physics Education*, vol. 56, no. 2, p. 025011, Jan. 2021. [Online]. Available: https://doi.org/10.1088/1361-6552/abce02

[12] H. Kim and G. Lee, “Reflection and outlook on special relativity education from the perspective of einstein: Focusing on research papers published in korea,” *Journal of the Korean Physical Society*, vol. 73, no. 4, pp. 422–435, Aug. 2018. [Online]. Available: https://doi.org/10.3938/jkps.73.422

[13] J. S. Aslanides and C. M. Savage, “Relativity concept inventory: Development, analysis, and results,” *Physical Review Special Topics - Physics Education Research*, vol. 9, no. 1, May 2013. [Online]. Available: https://doi.org/10.1103/physrevstper.9.010118

[14] M. Pitts, G. Venville, D. Blair, and M. Zadnik, “An exploratory study to investigate the impact of an enrichment program on aspects of einsteinian physics on year 6 students,” *Research in Science Education*, vol. 44, no. 3, pp. 363–388, Oct. 2013. [Online]. Available: https://doi.org/10.1007/s11165-013-9386-6

[15] T. Kaur, D. Blair, J. Moschilla, W. Stannard, and M. Zadnik, “Teaching einsteinian physics at schools: part 1, models and analogies for relativity,” *Physics Education*, vol. 52, no. 6, p. 065012, Sep. 2017. [Online]. Available: https://doi.org/10.1088/1361-6552/aa83e4

[16] R. K. Choudhary, A. Foppoli, T. Kaur, D. G. Blair, M. Zadnik, and R. Meagher, “Can a short intervention focused on gravitational waves and quantum physics improve students’ understanding and attitude?” *Physics Education*, vol. 53, no. 6, p. 065020, Oct. 2018. [Online]. Available: https://doi.org/10.1088/1361-6552/aae26a

[17] Y. S. Dua, D. G. Blair, T. Kaur, and R. K. Choudhary, “Can einstein’s theory of general relativity be taught to indonesian high school students?” *Jurnal Pendidikan IPA Indonesia*, vol. 9, no. 1, pp. 50–58, Mar. 2020. [Online]. Available: https://doi.org/10.15294/jpii.v9i1.22468

[18] M. L. Ruggiero, “Teaching einsteinian concepts in the first years of italian secondary school: some preliminary results,” in *The Fifteenth Marcel Grossmann Meeting*. WORLD SCIENTIFIC, Jun. 2022. [Online]. Available: https://doi.org/10.1142/9789811258251_0191

[19] M. Kersting, “General Relativity in Secondary School,” Ph.D. dissertation, University of Oslo, Oslo, Norway, 2019.

[20] A. Bandyopadhyay and A. Kumar, “Probing students’ understanding of some conceptual themes in general relativity,” *Physical Review Special Topics - Physics Education Research*, vol. 6, no. 2, Jul. 2010. [Online]. Available: https://doi.org/10.1103/physrevstper.6.020104

[21] ——, “Students' notions regarding ‘covariance’ of a physical theory,” *European Journal of Physics*, vol. 31, no. 6, pp. 1391–1413, Sep. 2010. [Online]. Available: https://doi.org/10.1088/0143-0807/31/6/008

[22] J. B. Hartle, “General relativity in the undergraduate physics curriculum,” *Am. J. Phys*, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0506075

[23] E. Baldy, “A new educational perspective for teaching gravity,” *International Journal of Science Education*, vol. 29, no. 14, pp. 1767–1788, Nov. 2007. [Online]. Available: https://doi.org/10.1080/09500690601083367

[24] T. R. Watkins, “Gravity & Einstein: Assessing the Rubber Sheet Analogy in Undergraduate Conceptual Physics,” Master’s thesis, Boise State University, Boise Idaho, 2014.

[25] R. Choudhary, A. Foppoli, T. Kaur, D. Blair, R. Burman, and M. Zadnik, “A comparison of short and long einsteinian physics intervention programmes in middle school,” *Research in Science Education*, vol. 52, no. 1, pp. 305–324, Jul. 2020. [Online]. Available: https://doi.org/10.1007/s11165-020-09944-8

[26] A. Postiglione and I. D. Angelis, “Students’ understanding of gravity using the rubber sheet analogy: an italian experience,” *Physics Education*, vol. 56, no. 2, p. 025020, Jan. 2021. [Online]. Available: https://doi.org/10.1088/1361-6552/abd1c4

[27] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, “Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis,” *Computers & Education*, vol. 70, pp. 29–40, Jan. 2014. [Online]. Available: https://doi.org/10.1016/j.compedu.2013.07.033

[28] C. J. Chen, S. C. Toh, and W. M. F. W. Ismail, “Are learning styles relevant to virtual reality?” *Journal of research on technology in education*, vol. 38, no. 2, pp. 123–141, 2005.

[29] J. Parong and R. E. Mayer, “Learning science in immersive virtual reality.” *Journal of Educational Psychology*, vol. 110, no. 6, pp. 785–797, Aug. 2018. [Online]. Available: https://doi.org/10.1037/edu0000241

[30] D. Weiskopf, M. Borchers, T. Ertl, M. Falk, O. Fechtig, R. Frank, F. Grave, A. King, U. Kraus, T. Muller, H.-P. Nollert, I. Mendez, H. Ruder, T. Schafhitzel, S. Schar, C. Zahn, and M. Zatloukal, “Explanatory and illustrative visualization of special and general relativity,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, pp. 522–534, 2006.

[31] E. McNitt and Novelab, “Spheres,” 2020. [Online]. Available: https://www.oculus.com/experiences/quest/3789736921099233/

[32] P. Lynch and University College Dublin Relativity Group, “Black Holes:Light & Matter,” 2020. [Online]. Available: https://sidequestvr.com/app/643/black-holes-light-matter

[33] A. Duffy and ARC Center of Excellence for Gravitational Wave Discovery, “Science in vr,” 2017. [Online]. Available: https://www.scivr.com.au/

[34] M. Kersting, R. Steier, and G. Venville, “Exploring participant engagement during an astrophysics virtual reality experience at a science festival,” *International Journal of Science Education, Part B*, vol. 11, no. 1, pp. 17–34, Dec. 2020. [Online]. Available: https://doi.org/10.1080/21548455.2020.1857458

[35] I. Radu and B. Schneider, “What can we learn from augmented reality (AR)?” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, May 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300774

[36] A. J. Magana, M. I. Serrano, and N. S. Rebello, “A sequenced multimodal learning approach to support students' development of conceptual learning,” *Journal of Computer Assisted Learning*, Mar. 2019. [Online]. Available: https://doi.org/10.1111/jcal.12356

[37] T. Kaur, D. Blair, R. K. Choudhary, Y. S. Dua, A. Foppoli, D. Treagust, and M. Zadnik, “Gender response to einsteinian physics interventions in school,” *Physics Education*, vol. 55, no. 3, p. 035029, Apr. 2020. [Online]. Available: https://doi.org/10.1088/1361-6552/ab764d

[38] S. Siegel and N. Castellan, *Nonparametric statistics for the behavioral sciences*, 2nd ed. McGraw–Hill, Inc., 1988.

[39] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50 – 60, 1947. [Online]. Available: https://doi.org/10.1214/aoms/1177730491

[40] W. H. Kruskal and W. A. Wallis, “Use of ranks in one-criterion variance analysis,” *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441

[41] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. [Online]. Available: http://www.jstor.org/stable/3001968

[42] E. Whitley and J. Ball, “Painless acute myocardial infarction and diabetes mellitus,” *Critical Care*, vol. 6, no. 6, p. 509, 2002. [Online]. Available: https://doi.org/10.1186/cc1820

[43] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, 1988 - 1988.

[44] L. V. Hedges, “Distribution theory for glass’s estimator of effect size and related estimators,” *Journal of Educational Statistics*, vol. 6, no. 2, pp. 107–128, 1981. [Online]. Available: http://www.jstor.org/stable/1164588

[45] R. R. Hake, “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” *American Journal of Physics*, vol. 66, no. 1, pp. 64–74, 1998. [Online]. Available: https://doi.org/10.1119/1.18809

[46] J. M. Nissen, R. M. Talbot, A. Nasim Thompson, and B. Van Dusen, “A comparison of Hake’s g and Cohen’s d for analyzing gains on concept inventories,” *arXiv e-prints*, p. arXiv:1612.09180, Dec. 2016.

[47] A. I. of Physics, “Percent of physics bachelor’s earned by women, classes of 1976 to 2020,” 2021. [Online]. Available: https://www.aip.org/statistics/data-graphics/percent-physics-bachelor%E2%80%99s-earned-women-classes-1976-2020

[48] R. Ilie, E. Shaffer, C. M. D’Angelo, D. Cermak, M.-Y. Lin, and H. Chen, “Virtual reality laboratory experiences for electricity and magnetism courses,” in *2021 ASEE Virtual Annual Conference Content Access*. Virtual Conference: ASEE Conferences, July 2021, https://peer.asee.org/38025.

[49] J. P. Canright and S. W. Brahmia, “Developing expertlike epistemologies about physics empirical discovery using virtual reality,” in *Proceedings of the Physics Education Research Conference (PERC*, M. Bennett, B. Frank, and R. Vieyra, Eds., Oct. 2021, pp. 75–80.

[50] J. R. Brown, I. Kuznetcova, E. K. Andersen, N. H. Abbott, D. M. Grzybowski, , and C. D. Porter, “Full paper: Implementing classroom-scale virtual reality into a freshman engineering visuospatial skills course,” in *2019 FYEE Conference*. Penn State University , Pennsylvania: ASEE Conferences, July 2019, https://peer.asee.org/33701.

[51] D. Iquira Becerra, J. Herrera Quispe, R. Aceituno, G. Vargas, F. Zamora, J. Mango, G. Anccasi Figueroa, A. Vizcarra, and J. Chana, “Evaluation of a gamified 3d virtual reality system to enhance the understanding of movement in physics,” in *Proceedings of the 9th International Conference on Computer Supported Education - Volume 1: CSEDU*, 01 2017, pp. 395–401.

[52] D. McGrath, C. Savage, M. Williamson, M. Wegener, and T. McIntyre, “Teaching special relativity using virtual reality,” in *Proceedings of the Assessment in Science Teaching and Learning*, 01 2008.

[53] E. Poisson and C. M. Will, *Gravity*. Cambridge University Press, May 2014. [Online]. Available: https://doi.org/10.1017/cbo9781139507486

Appendix

A The Physics of the Virtual Reality Demonstration

Here we discuss the physical equations implemented to curve space and dilate time within the virtual reality (VR) simulation of this study. This section is intended for an audience familiar with general relativity (GR). The background and derivations of the fundamental concepts discussed can be found in any introductory GR textbook. The reader should note that the focus of this section is not on mathematical rigor but rather on formulating expressions derived from GR optimized for practical implementation and visual understanding in VR.

For our demonstration, we are interested in the curvature of spacetime outside a spherically symmetric non-rotating mass. This is given in its exact form by the Schwarzschild metric, derived from Einstein’s field equations. The line element associated with the Schwarzschild metric in natural units takes the form [53]

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{1}{1 - \frac{2m}{r}} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{3}$$

where dr, $d\theta$, and $d\phi$ are infinitesimal spacetime intervals as measured by an observer far from an object of mass m. Together, these provide an expression for the invariant spacetime interval ds as measured by an observer a distance r from the gravitating body.

To decouple the curvature of space from that of time for visual representation in three dimensions, we slice our four-dimensional spacetime through a moment in time so that t is constant. Furthermore, taking advantage of the spherical symmetry of the scenario, we consider a slice through the equatorial plane of the massive object with $\theta = \pi/2$. Additionally, we narrow our scope to the change in radial distances between events along a line of constant ϕ, and Eq. (3) reduces to

$$d\sigma^2 = \frac{1}{1 - \frac{2m}{r}} dr^2, \tag{4}$$

where $d\sigma$ is the infinitesimal radial length component measured by an observer at their location r, near the gravitating body.

Thus, the factor by which the radial length component at radial distance r near the mass as seen from the perspective of an observer at infinity differs from the same length component as seen by an observer at radial distance r is given by

$$\frac{dr}{d\sigma} = \sqrt{1 - \frac{2m}{r}}. \tag{5}$$

In the simulation, this equation is used as a scaling factor for the displacement of the grid point positions as a function of their proximity to the center of the mass.

Similarly, the equation for time dilation considering a constant spatial location (i.e. r, θ, and ϕ are constant) follows directly from Eq. (3) with

$$d\tau^2 = -ds^2 = \left(1 - \frac{2m}{r}\right) dt^2, \tag{6}$$

where $d\tau$ is an infinitesimal time interval measured by the clock of an observer at r, close to the massive object, and dt is the time interval measured by a distant observer watching the same clock positioned at r. Thus we find that the factor by which a time interval at r measured by an observer at infinity differs from the same interval at r measured by an observer at r is given by

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{2m}{r}}}. \tag{7}$$

Within the simulation, the inverse of this factor is used to calculate the change in frequency of the hands of the visual clocks as the mass is moved towards and away from them.

B Survey Instruments

Throughout this section we list all of our survey items. We begin with those questions that are only on the pre-survey, then list questions common to both the pre-survey and post-survey, and finish with questions that are only on the post-survey. For all of the multiple choice questions, the answer choices are listed with the correct answer highlighted in bold, and are followed by a confidence rating (listed only after Q1 for brevity).

Pre-survey only:

Q: I am in my (...) year of college.

- First
- Second
- Third
- Fourth
- Fifth+

Q: Please list your major:

- _____

Q: What is your gender identity?

- Male
- Female
- Non-binary
- Prefer to self-identify _____
- Prefer not to say

Q: Have you used a virtual reality headset before (i.e. Google Cardboard, Oculus headset, HTC Vive)?

- Yes
- No

Q: Please rate your level of prior experience with virtual reality.

- No experience at all
- Little experience
- Some experience
- Regular experience
- Extensive experience

Q: Please rate your level of prior experience with general relativity.

- No experience at all
- Little experience
- Some experience
- Regular experience
- Extensive experience

Both pre-survey and post-survey:

Q1: Is the curvature of spacetime the same on all sides of a spherical object?

- **Yes**
- No

Confidence Q: How confident are you about your answer?

- Guessing
- Hesitant
- Neutral
- Confident
- Certain

Q2: Does the amount of spacetime curvature relate to gravitational field strength?

- **Yes**
- No

Q3: How does a gravitational field affect the passage of time?

- **Time passes more slowly in a gravitational field.**
- Time passes more quickly in a gravitational field.
- Time does not change in the presence of a gravitational field.

Q4: A more massive object has _____ effect on the gravitational field.

- **A greater**
- A lesser
- The same

Q5: When you release a smaller mass in the presence of a larger mass, which direction does it fall?

- Down
- **Towards the center of the larger mass**
- Both

Q6: Imagine two objects with different masses being held apart in space. When they are released, what happens?

- **They both move**
- Only the smaller mass object moves
- Only the larger mass object moves

Q7: In the following set of diagrams there are two separate objects with different masses. The object on the left is much more massive than the object on the right. Please select the set of diagrams that correctly illustrates the relative curvature around each object.

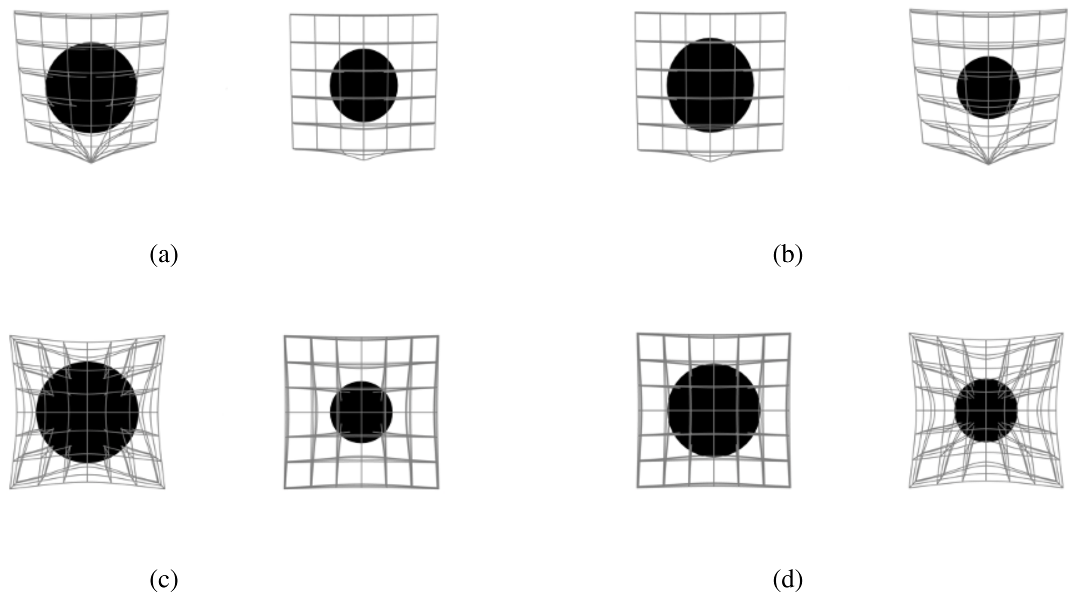


Figure 4: Answer choices for Q7; the correct answer choice is (c).

Q: Please rate your agreement with each of the following statements. (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree)

- Statement: I think physics is interesting.
- Statement: I enjoy learning new scientific concepts.
- Statement: I prefer to learn physics through activities rather than through lectures.
- Statement: General relativity is an important idea for modern science.
- Statement: I want to learn more about general relativity.
- Statement: I can understand new physics topics quickly.
- Statement: I could be a good physicist.
- Statement: Science is only for smart people.
- Statement: Physics is useful and relevant to me

Post-survey only:

Q: In your own words, please explain how space, time, and gravity are connected.

- ______

Q: I participated in the...

- immersive virtual reality demonstration
- tabletop demonstration

Q: Please rate your agreement with each of the following statements. (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree)

- Statement: I liked the graphics/visuals within the activity.
- Statement: I continued the activity because I was curious.

- Statement: I was engaged throughout the activity.
- Statement: I was interested in the activity.
- Statement: I thought the activity was worthwhile.
- Statement: I thought the equipment was easy to use.

Q: Please leave any additional comments you have about the usability of the activity.

- _____