Comparison of the Effect of Interactive versus Passive Virtual Reality Learning Activities in Evoking and Sustaining Conceptual Change

Maria Roussou and Mel Slater

Abstract—The key question we address is whether interactivity in a virtual environment (VE) impacts conceptual learning. We developed a Virtual Reality (VR) application to simulate a playground, in which children had to engage in tasks that required solving arithmetical fractions problems. Fifty (50) children were tested in an empirical study with different conditions that varied the levels of interactivity and immersion, from a fully interactive to a non-immersive and non-interactive activity with LEGOs. A methodological framework, where quantitative and qualitative analyses complement each other, was developed to analyze children's activity. Our findings from the quantitative analysis showed that participants in the VEs seemed to have a greater overall gain than those that did not perform the activity in VR but did not clarify whether interactivity in the VE was the defining factor in this gain. The qualitative analysis indicated that it was the "passive" VR condition which provided evidence of sustained conceptual change. With countless educational VR applications becoming available nowadays to the public due to VR's recent resurgence in prominence through inexpensive headsets, the study of the effect of VE attributes, such as interactivity, is timely.

Index Terms—Educational technology, Human computer interaction, Interactivity, Virtual Reality

1 Introduction

INTERACTIVITY has played a core role throughout the evolution of educational technology, itself evolving from the initial clicking on a button to advance through a set sequence required by programmed instruction, to permitting users to create the educational content themselves. Our research is based on the notion that interaction within an immersive Virtual Reality (VR) environment, in which children can reach out and transform objects in a surrounding 3D world, can enhance learning of abstract concepts such as the understanding of arithmetical fractions. VR in education has been used before for subjects that naturally lend themselves to visual display [20]. However, in our case the idea is to apply VR to more abstract problems, such as the learning of mathematics at early school age (8-12 years). Interaction within a virtual environment (VE) is qualitatively different from interaction in a desktop system. In the former there is the capability for the mobilization of the participant's whole body in the task, e.g. reaching out and touching something, in order to grab and move it. This is completely different from the use of, say, a mouse to 'touch' an object on a 2D display. The question we asked is whether such interactivity can make a difference in the learning process. The answer is probably 'yes' for tasks that naturally lend themselves to this representation – such as actual motor tasks or the learning of a manual skill. However, would the same apply to more abstract learning? Is the almost physical manipulation of objects that VR makes possible helpful in the problem of learning?

- *Maria Roussou is with the Department of Informatics and Telecommunications, University of Athens, Panepistimiopoli, Ilissia, GR 157 03, Athens Greece. E-mail: mroussou@di.uoa.gr.*
- *Mel Slater is with the Department of Psychology and Psychobiology, University of Barcelona And Institució Catalana de Recerca i Estudia Avançats (ICREA), Campus de Mundet – Edifici Teatre, Passeig de la Vall d'Hebron 171, 08035 Barcelona Spain. E-mail: melslater@ub.edu*

Our work has, thus, focused on investigating interactivity in immersive virtual learning environments and its effect on conceptual learning, through the design of a set of environments and experiments to evaluate children's behavior in VR [29]. Empirical studies were carried out with a total of 60 primary school students (ages 8-12), in a number of different experiments: an exploratory study was carried out with 10 children to test the methodology and prepare for the main study [30]; the main study, a large scale experiment described and discussed in this paper, was conducted with 50 children who played with different virtual and physical environments that simulated a playground, which focused on a presentation of problems in mathematical fractions. Three conditions - an interactive VR, a "passive" (or guided) VR, and a non-VR condition using LEGO bricks - each with different levels of interactivity, were designed to evaluate how children accomplish the various conceptual tasks. Pre-tests, post-tests, interviews, video, and computer logs were collected for each participant, and analyzed both quantitatively and qualitatively. Children's responses to the test questions were analyzed quantitatively using logistic regression. Qualitatively, the descriptive framework of Activity Theory (AT) was used to analyze user behavior in the immersive VEs and to identify conceptual contradictions, i.e. the occurrence of critical incidents, focus shifts or breaks in the elements of the learner's activity that led to indications of the learner's construction of meaning [21].

Next, we describe how we address the gap identified in related research, followed by a presentation of the main empirical study carried out, and the analyses of the resulting data. Finally, we discuss our experimental findings and directions for future work.

xxxx-xxxx/0x/$xx.00 © 201x IEEE Published by the IEEE Computer Society

2 Background and Related Work

The digital revolution is witnessing an increasing number of projects, products, and technology-enhanced play environments that are directed to young students, across a wide spectrum of applications, from microworlds to immersive virtual worlds. Interactivity has been, in different ways, the driving force behind these efforts. We consider interactive learning environments from a number of different contexts (informal education, research, etc.), focusing on learning environments that use VR as their main display medium.

The informal education context, for instance, has seen a few such projects, most of which have been developed with a focus on "edutainment" (e.g. [25], etc.). Interactivity plays a key role in these types of digital environments where the learning goal is more free-form and exploratory, enabling the occurrence of "unexpected events, novel reactions, novel activities, and novel combinations of activities or events, which in turn facilitate children to question them and to reflect on their experiences" [27]. However, in most immersive VR environments for children a non-critical position toward interactivity is assumed as the emphasis is on the immersive and representational aspects of the simulations rather than the interaction methods and modalities. Thus, the dominant model of interaction in VEs targeted for public presentation has been the exploration of virtual space through navigation. The difficulty, on both a technical and a conceptual design level, of incorporating other forms of interaction has prevented designers of education spaces from exploring more exciting and innovative models of communication with the virtual world. Interactivity is largely restricted and difficult to apply in a public space, especially when the practical difficulties of visitor throughput and other complications must be overcome or when more than one user must share the same screen [26]. Additionally, very few of these projects have carried out any research on their actual impact, and where such studies exist, these are mostly focused on technical descriptions of the systems and usability rather than on studying any effects of the environments on learning.

In the research domain, a core of high quality research work and theoretical frameworks emerged in the mid-nineties, making the VR medium a legitimate arena in which to develop and test applications for educational purposes [10], [36], [20], [28]. More recently, interactivity has been explored in the context of tangible, embodied and Augmented Reality environments [14]. Nevertheless, despite the drive to empirically prove the value of VR for education, both quantitative and qualitative studies performed for most of the developed projects have not been able to report much with respect to children's conceptual learning in the VEs, although some reported achievement of the learning goals they had initially set out to achieve. Again, the effect of interactivity has been debated [15], as has its benefit to learning [23].

The connection that binds interactivity and learning seems to be strong and there are reasons to believe that interactivity may be a defining component in the successful outcome of a virtual learning environment. Nevertheless, the lack of studies that explore the role of interactivity in VR-based learning leaves a gap in the development of research that calls for further examination. The research questions that have emerged ask if and how interactivity in a VE can influence learning. Next, the main study conducted to answer these questions is presented.

3 Main Experiment on Learning Fractions

A study was carried out with 50 primary school students (ages 8-12), in a between-groups experiment with three conditions. The experiment was conducted with two different environments designed to simulate a playground: an immersive VE entitled the 'Virtual Playground' (VP) and a physical model constructed with LEGO bricks.

In both cases, participants were asked to complete a set of tasks designed to address arithmetical 'fractions' problems, i.e. tasks with which young students tend to have deep conceptual difficulties [7]. The three conditions, two based on the Virtual Playground and one with the LEGO, were designed with varied levels of interaction and form of activity. The first experimental condition in the Virtual Playground was an interactive VR condition (IVR) where children had full control over the virtual playground in a CAVE-like installation, wearing a head tracker and using a 3D wireless mouse to perform the tasks (Fig. 1). The other Virtual Playground condition was a 'passive' VR condition (PVR) where children, again in the same CAVE, observed a robot (named 'Spike') playing out the tasks in a pre-recorded sequence of actions. Finally, the non-VR condition was carried out with children who used the physical model to design a playground with LEGO bricks.

Fig. 1 A 9 y. old girl using the wand to pick and place blocks in the interactive version of the Virtual Playground (IVR).

3.1 The Virtual Environment and the Design of Interactivity

The Virtual Playground is a highly immersive VE, created for projection-based VR systems and tailored to provide an evaluation tool for young user behaviour in VR. The playground includes various elements such as swings, a slide, a roundabout, etc., which are incorrectly scaled and positioned. It presents an interesting challenge in that it requires the participant's physical and mental intervention in order for the elements in the environment to be correctly scaled and positioned. The participant's task is to change the size of the area that each object occupies in the playground according to rules that are provided during the experience (e.g. "The slide is smaller than it should be; its area must be increased by 1/5 of the area it covers now" is spoken out by a bird character floating above the slide area). These rules require the participant to make fractions calculations; however the task is presented in the form of a game, with a narrative structure including engaging visuals and characters (Fig. 2), in an attempt to disguise the maths and any potentially instructional format.

Fig. 2 An owl introduces the story and overall goal to the participant (left); A virtual "talking bird" floats over each area of the playground that needs to be changed, speaking out the rule for that area (right).

Drawing from Pares and Pares' classification of explorative, manipulative, and contributive forms of interaction [24], an effort was made to include elements from all three of these forms of interaction in the environment of the Virtual Playground. Specifically, e*xplorative interaction* is supported through the ability to navigate freely in the environment, the lowest possible form of interactivity in the VP. *Manipulative interaction* is supported by the ability to pick and place elements in the environment or switch points of view to review things. In the VP, this kind of interaction is the main form of activity on an operational level. Finally, *contributive interaction*, which refers to the most involved form of interaction, is the ability to alter the system itself. The VP supports contributive interaction, since the participant is able to change the environment as a whole (i.e. the playground being different when the participant comes out of it from when she went in). What is of interest to this research, however, is contributive interaction that takes place on a conceptual level; in other words, interaction that leads to conceptual change or contributes towards a change of the child's mental models.

In order to support this kind of interaction, in addition to the aforementioned rules, constraints and a related system feedback mechanism have been designed into the VP, aiming at establishing reciprocal activity between the participant and the system. System feedback provides constraints on where the blocks can be placed with a purpose to constrain participants' actions in such a way as to provide conceptual conflicts (or contradictions, to use the Activity Theory terminology described later in the paper), especially for students that have been struggling with the content in the first place. Another goal of the system feedback has been to minimize external (human) instruction or intervention, e.g. feedback by the observer.

System feedback in the VP has been programmed to *inform*, *confirm*, or *react* to incorrect participant activity by *prompting* for further action. For example, feedback to inform is presented to the child by the virtual owl, while the rule for each area is provided by a same-colored bird, which floats over that area and talks to the participant when clicked upon. Confirmation feedback is provided through sound effects that confirm each action and through verbal statements such as "This is too close to the fence". The system provides reactive feedback concerning placement of the blocks onto the playground tiles, since not all tiles of the playground are available for placing blocks on. For example, if the tile where a block is to be placed is adjacent to the playground fence or a bench, or is any one of the other "forbidden" tiles (i.e. footpath tile, entrance gate tile, tiles around the pool), then the block cannot be placed and a different kind of system feedback sound for each case will be heard. Additionally, each block, when placed on a tile on the grid, is checked against every adjacent tile on all four sides. If any of the adjacent tiles are of the same type (i.e. of the same color), then the block can be placed. If not, then the system does not allow its placement and responds to the participant with appropriate audio feedback. For usability reasons, each block, when placed on the grid, adjusts and snaps into place to facilitate the child's physical task and leave room for more concentration on the conceptual task. In most cases, visual and audio cues are given to enhance the constraints and restrictions designed into the environment.

In the case of the passive VR experience, the user is immersed in the VE but does not have the ability to interact with it, i.e., to explore freely via navigation, manipulate objects or receive system feedback through the mechanisms described earlier. Rather, interaction takes place between the system and the virtual robot, which the participant can observe as in passively watching an "immersive video". For the purposes of this research, the choice of events and activity that were recorded followed a long testing process based on observing typical participant interaction with the interactive version of the Virtual Playground during the pilot study. An expert person's activity in the VE was then recorded, including all movement, manipulation and visual and audio feedback received by the person from the system at the time of interaction. As a result, the recorded sequence was timed to include use of all the features provided by the interactive VE, the successful completion of all the tasks, and the necessary pauses, within 27 minutes. Finally, in the LEGO condition, the goal of the activity was the same but instructions and rules were written on colored cards while no feedback was given to users. For a detailed description of the VP, see [29] and the website[1].

3.2 Procedure

Each session was conducted with one participant at a time and the duration of the session was approximately 2 hours for each child. Each participant in the VR conditions was accompanied to the study location by a parent or guardian, who was asked to complete an informed consent form as part of the ethics committee requirements. The participant was then asked to, first, fill out a questionnaire de-

[1] For images & video: https://www.makebelieve.gr/mroussou/vp

termining demographic information and level of prior experience with computers, game playing, and VR, and then a questionnaire with maths questions that were based on the fractions questions found in standardised tests (such as the UK Key Stage 2 SAT maths and science test). The questionnaire sought to reveal students' current understanding of fractions, and included a total of 11 questions (9 plus 2 questions with 2 parts) which covered different kinds of maths problems. The questions involved general fractions questions, fraction comparisons, and questions about fractions represented in a variety of ways. The latter set of questions was included to test the participants' ability to translate between different representations of fractions.

Once these were completed, the participant took part in one of the experimental conditions (a between-groups design). Participant's pre-test scores (but also the collection of information about each student's performance from the teacher or parents) were used to derive aptitude data (previous knowledge of fractions) and thus classify them in aptitude groups: low aptitude (from 0 to 6 correct responses), medium (from 7 to 9 correct responses) or high (10 or 11 correct responses). This information was then used to distribute the children to the experimental condition accordingly, so as to ensure that the aptitude distribution of participants across conditions was balanced. The rationale behind the even distribution of participants of different levels across conditions was to avoid assigning students of high aptitude into the same group, as well as to avoid a ceiling effect, where advanced participants would find the exercises to be too easy. Aptitude treatment could also facilitate the analysis of questions such as if the VR experience helped good students more, less, or made no difference, or, if boys who performed better than girls in the pre-test, also did better when using the VE and vice versa, etc.

For all three experimental conditions, the observer was constantly present, observing, asking questions and encouraging the participant to explain her/his actions while doing (by thinking aloud). The same level of observer "guidance" was given to all, ensured by the fact that the observer was the same person for all 50; this was one of the reasons that the experiments took months to conclude. The difficulty in the role of the observer had been recognized prior to the study. Many researchers [6], [22], [13], [19] have noted that the observer must be as unobtrusive or as uniformly intrusive as possible, while ensuring that the participants are occasionally reminded to think-aloud. There is always the risk of the observer intervening too much to help the participant, possibly altering the result. We carefully chose a variation of the think-aloud technique called 'Active Intervention', where the observer prompted the children for explanations of what they were doing and to give a commentary on their interaction. Usability evaluation sessions with children carried out by [16] have shown that most verbal comments were gathered during Active Intervention sessions versus other methods such as the think aloud or co-discovery, and that quiet children are better able to provide verbal comments when a more active way of prompting is applied, for example by asking questions. Hence, for these studies, it was decided that the observer would assume the more active role of a "cultural informant" rather than that of a "detector" who passively records what is "there" [1]. Nevertheless, a limitation of our studies is that no post hoc analysis was performed to assess the human experimenter's impact on our results.

After the experience, a post-test had to be completed. This was identical to the pre-test in terms of the underlying conceptual requirements, which means that the only differences between pre- and post-test were in the numbers used for the fractions. Finally, each participant was interviewed for a duration of 10 - 15 minutes in a conversational semi-structured format by the observer, who noted the specific actions which the participant had problems with, and directed the participant to reflect on these accordingly.

An effort was made to keep the duration of the pre-test, the actual task, and the post-test under 30 minutes each. For participants in the VR conditions, short breaks were encouraged throughout the experience (at 10 - 15 minute intervals) to prevent any possible effects of simulator sickness and discomfort (however, most participants did not wish to take breaks). Overall, the duration of the experiment for each, including training, pre-test, main task, post-test, and interview did not exceed 2 hours, while the non-VR condition was usually significantly shorter.

3.3 Participants

Data from a total of 50 (N=50) participants was acquired, 17 (8 boys and 9 girls) from the IVR condition, 14 (9 boys and 5 girls) from the PVR condition, and 19 (8 boys and 11 girls) from the LEGO condition. Overall, a balanced distribution of gender was achieved, with a total of 25 male and 25 female participants for all conditions. However, the distribution of male and female participants was not equal within each condition, due to the difficulties encountered in recruiting the sample. The mean age of the participants was 10.4 years. Nevertheless, since age is not always indicative of aptitude, the age range of 8-12 was specified mostly as a starting point and further questions were asked during the recruitment process in order to identify whether each participant fit the specific criteria placed by the study. In other words, it became apparent at the outset that the understanding of fractions varied greatly between participants of the same age, so age could not be the determining factor for finding children that had been taught fractions in school. Finally, it was each student's ability (as judged from the results of the pre-test and information collected from parents) that became the defining factor of the student's level of knowledge when coming into the study.

4 Analysis of Main Study

The focus of the main study was to capture instances of learning that were triggered by interactive activity in the VE and which appeared to lead to conceptual understanding of fractions. To identify these instances of conceptual learning a number of measures were taken to ensure that the data collected would result in a wealth of information, which could be meaningfully combined and analysed. As [6] argues, the analysis of data is only as good as the way the data was collected. Thus, multiple different methods of gathering information were designed, ranging from the quantifiable pre- and post- questionnaires and computer log files to the more qualitative direct observations and interviews. The particular idiosyncrasies of working with children were taken into account (such as their varied ability to verbalise and limited attention spans) and, as a result, the most appropriate methods were formulated [22]. The data collected through these methods were analysed quantitatively and qualitatively.

4.1 Quantitative Analysis

The quantitative analysis was applied to the data collected through the questionnaires. Two response variables were considered in the analysis; *postcorrect*, which corresponds to the number of questions correctly answered after the experiment and *postattempt*, which is the number of questions attempted after the experiment. The independent variable was the condition: Interactive VR (IVR), Passive VR (PVR), or the non-VR condition (LEGO). The explanatory variables were: gender ('*gender*'), age of participant ('*age*'), year in school ('*yearinschool*'), frequency of computer usage ('*computer*'), computer game playing ('*gameplay*'), and the number of correct responses on the pre-test ('*precorrect*'). Since the response variables can be treated as binomially distributed random variables, they were analysed using the logistic regression method, i.e. the number of correct answers (r) out of the number of questions (n) would have a binomial distribution [37].

The variables were constructed from the $n=11$ questions of the pre-test and post-test. The number of correct responses was counted out of all responses and the response variables were related to linear combinations of the independent and the explanatory variables.

TABLE 1
MEAN±STANDARD DEVIATIONS OF CORRECT RESPONSES IN THE PRE-TEST AND POST-TEST FOR EACH OF THE 3 CONDITIONS.

Condition	Precorrect Mean±StdDev	Postcorrect Mean±StdDev
IVR (N=17)	6.6±3.1	7.2±3.7
PVR (N=14)	6.4±3.9	7.1±1.9
LEGO (N=19)	5.3±2.4	6.5±4.3

4.1.1 Analysis for the 'postcorrect' response variable

The variable *postcorrect* is the number of questions correctly answered after the experiment. All questions that were not answered (not attempted) were considered incorrect under the assumption that if participants, for whatever reason, did not attempt a question then they were not able to answer it correctly. The means and standard deviations of the *postcorrect* response variable for each of the three conditions are shown in Table 1. Logistic regression was used to model the relationship between the dichotomous *postcorrect* response variable and the explanatory variables. The significant variables that were fitted into the overall model, with the change in deviance (χ^2), the degrees of freedom (d.f.), the direction of the association, and the p-value are depicted in Table 2. The table lists only the variables that had a significant impact on *postcorrect*. The chi-squared (χ^2) values in the table indicate the increase in deviance that would result if the corresponding variable were eliminated. Each calculated χ^2 value is compared with the tabulated χ^2 5%[2] value, which is 3.84 on 1 d.f. and 5.99 on 2 d.f. If the change in deviance caused by deleting a variable from the fitted model is greater than the tabulated chi squared 5% value then the variable is significant.

TABLE 2
FITTED LOGISTIC REGRESSION FOR THE 'POSTCORRECT' RESPONSE VARIABLE.

Variable	change in deviance (χ^2)	d.f.	association	p-value
precorrect	93.96	1	positive	0
conditionPVR. precorrect	8.02	2		0.0181
condition-LEGO.precorrect			negative	
yearinschool	4.26	1	positive	0.0390
overall	70.61	42		

The condition IVR, PVR, or LEGO, did not impact *postcorrect*. On the other hand, independently, *precorrect* was the most significant variable associated with *postcorrect* (χ^2= 93.96>>3.84, d.f.=1), as would be expected. However, there was a significant interaction effect between condition and *precorrect* (*conditionPVR.precorrect* χ^2 = 8.02>5.99, d.f.=2). Therefore, given the same starting level (*precorrect*), there is greater gain amongst participants in the VR conditions (IVR and PVR) than amongst the LEGO participants (Fig.3). It is, however, notable that the latter's *precorrect* scores were worse; hence the use of the *postcorrect* variable is important to remove the effects of difference in aptitude.

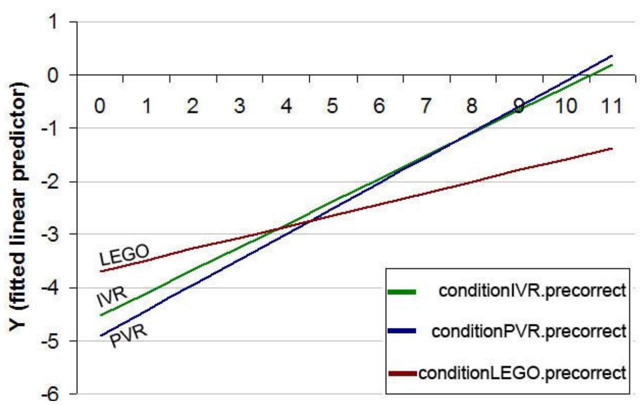

Fig. 3 Linear predictor for postcorrect (vertical axis) against precorrect (horizontal axis) under the 3 experimental conditions.

The *yearinschool* also impacts *postcorrect*. Finally, the overall deviance is 70.611 on 42 d.f. In order for the model

[2]These can be retrieved from a standard χ^2 table or a MATLAB function.

to be of a good fit at the 5% level of significance, the overall deviance should be less than 58.125 on 42 d.f. Therefore, although the variables making up the model for the *post-correct* variable are significant, the fit could be significantly improved by the addition of other variables that were not recorded in this study.

In summary, the model that fit the data for the *postcorrect* response variable showed the following:

- *precorrect* is associated with *postcorrect*, as would be expected. This is a strong positive association and, by far, the most significant result. This variable is important because it removes the effect of the participants' aptitude level when they came into the experiment.
- the slope of the regression for *postcorrect* against *precorrect* is not different between the IVR and the PVR conditions (*conditionPVR.precorrect* is not significantly different from *condition-IVR.precorrect*). The slope of the line is, however, lower for the LEGO condition (*condition-LEGO.precorrect* is significantly different from *conditionIVR.precorrect*). The overall increase in deviance for removing the interaction term is significant at about 0.02. That is, the two VR conditions give, on the average, a higher *postcorrect* score for the same *precorrect* score compared to the non-VR condition.
- a strong positive association between *yearinschool* and *postcorrect*, meaning that the number of years in school is significant and positively associated with the number of *postcorrect* responses, as would be expected.

The same strategy as above was used to analyse the *postattempt* response variable, i.e. the number of questions attempted (whether right or wrong), and resulted in the following (which is valid only for the IVR and PVR conditions since the LEGO condition results in 11/11 scores):

- a strong positive association with *preattempt*
- a positive association with *yearinschool*
- the slope of the relationship between *preattempt* and *postattempt* is lower for the PVR than for the IVR condition. However, this is a 'weak' relationship, i.e. significant at 10% but not at 5%.

4.1.2 Other variables

In the logistic regression analysis, *gender* was one of the variables shown to have no effect on either *postcorrect* or *postattempt*. Participants' gender was examined due to an expectation, based on public perception, that boys are better in maths and on studies indicating that boys' early and sustained exposure to video gaming places them at an advantage with respect to competence in technological environments [5]. However, our analysis did not confirm this.

Similarly, frequent computer usage and game playing could be thought of as providing an advantage to those that used the VE but were not found to have an effect on the number of correct or of attempted post-test responses.

We also analysed the computer log file, which recorded, among others, the time it took for each participant in the IVR condition to complete the activity. There was a significant negative correlation between pre-test performance and time taken, with students who did worse taking longer to complete the tasks in the VE. No correlation was found between the time taken to complete the playground in IVR and other variables. The duration of the PVR experience was pre-programmed to 27 minutes.

4.1.2.1 Performance on individual questions.

In addition to comparing overall pre- and post-test scores per condition, it was considered important to examine separately the outcome on two specific questions, Question A5 and Question A2. Question A5, asked both on the pre-test and the post-test, dealt with comparison of two fractions. There were two parts to the question: one asked to compare fractions of the same numerator ("Circle the fraction that is larger between 1/3 and 1/4"), and the second asked to compare fractions of a different numerator and denominator (such as 2/3 and 1/4) and circle the one that is smaller. The reason for focusing on this question is the well-known difficulty that children have with comparing fractions, as noted in the literature [3] and confirmed by our previously held exploratory study. Hence, a comparison of pre- and post-tests on this question was used to assess the effect of the study on the student's ability to compare fractions. The result indicates that the study did not have a significant effect on the ability to compare fractions which would be worth mentioning.

On the pre-test, Question A2 asked to find 1/6 of twelve ("Jack's father cuts a birthday cake into 12 pieces. He is going to give 1/6 of the cake to Jack and keep the rest for himself. How many pieces of cake will he give to Jack?"). In the playground activity, a similar problem was posed for the monkey bars, which had to be reduced by one sixth of the area of the sandpit, i.e. 1/6 of 12. Another similar task in the playground involved the slide, which required that its area be increased by one fifth of ten. During the activity in the VE or with LEGO, a number of participants seemed to have difficulty in answering this question. The tendency was to use the denominator of the fraction as the result (e.g. answering 6 for 1/6 of 12 or 5 for 1/5 of 10).

A comparison of pre- and post-tests on this question was used to assess the effect of the study on the student's ability to solve this problem successfully. However, as with the previous question, no significant differences were found that are worth mentioning.

4.1.3 Discussion of quantitative analysis results

The most significant associations found by the quantitative analysis implied that participants in the VEs had a greater overall gain than those that did not perform the activity in VR. Nevertheless, these results were not useful predictors of either performance or -better yet- improvement on post-test scores due to the existence or lack of interactivity. Moreover, they do not reveal details about what the participant has learned and how, and if the various levels of interactivity or types of system feedback given within the IVR condition are contributing. In other words, this analysis was able to provide some evidence with respect to the results on the post-test scores and post-test attempts but not a full response to the basic research question, i.e. whether interactivity in a virtual environment has any effect on children's ability to learn fractions.

As a result, two possible reasons can be identified; that either interactivity has no role to play in supporting learning or that quantitative analysis, on its own, is unable to fully identify this role. Some researchers have recognised that by focusing on quantitative techniques some important parts of a situation can be missed, and that often the best results are achieved through the use of mixed method evaluations where quantitative and qualitative analyses can complement each other [31]. Hence, the qualitative analysis described in the next section was conducted to throw further light on the results of the quantitative analysis and to capture the dynamic nature of the process of interaction in a VE and its effect on learning.

4.2 Qualitative Analysis

The experiments resulted in an enormous pool of data from multiple data sources, mainly video and audio recordings of the participants' activity during the actual experiences designed for the main study.

Upon close examination of the pre- and post-tests and the observation transcripts, some emerging themes were identified in terms of the learning content, related to the tasks, which proved to be difficult for many of the participants. The problems that more children seemed to have difficulty with were concentrated on issues such as fractions comparison (Question A5) and the tasks that related to Question A2 of the tests. These problems were identified on the basis of the frequency with which they occurred in all three of the conditions of the study (IVR, PVR, and LEGO), they were categorised into themes based on the contradictions or focus shifts that they illustrate, and analysed using an Activity Theory framework.

AT was adopted and adapted as one of many different approaches and theories to describe human interaction with machines (a comprehensive survey can be found in [27]). Bødker [21] introduced AT to the HCI community in an attempt to provide an alternative framework to cognitive science [4]. According to the extended systemic model of AT [12], an activity is composed of a **Subject** (a person or group engaged in an activity) and an **Object-O** (for instance, a learning objective held by the subject), mediated by a **Tool–T** or tools (that could be material as well as mental) as well as **Rules–R** (that regulate actions and interactions), the **Community-C** (one or more people who share the objective with the subject), and the **Division of Labour -DofL** (how tasks are divided between cooperating members of the community as well as the division of power and status), thus forming the complete Activity System. The systemic and dialectic nature of the complex and constantly evolving interrelationships between the above are illustrated by "triangle" diagrams (Fig. 4).

Hence, in AT terminology, each problem or task in the playground activity becomes an Object with an immediate, defined goal (e.g. the Object to solve 1/5 of 10 in order to correct the slide area). All Objects are part of the overall learning goal or Outcome, which is to understand fractions. Understanding fractions is an implicit outcome for the participant; on the other hand, to design a correct playground is the explicit goal. This analysis is interested in examining whether the implicit outcome of understanding fractions takes place and to identify the role of interactivity, as a property of the Tool (VR), in achieving this outcome. For the purpose of this analysis, the tasks have been grouped thematically to form categories of fractions problems where interesting contradictions were observed or where cues/features of the system prompted participants to act or respond in similar ways. It is the working through contradictions or the focus shifts (deliberate changes of focus of activity) [4] that are analysed and that, according to AT, are sources of development [17].

4.2.1 Thematic categories of problems

There are various aspects of this research problem that make AT an appropriate methodology for studying it, including the novelty and context of the tool (a VE), the richness of human activity in relation to it, and the dynamic nature of the goal of the activity (to achieve conceptual change). AT allows explaining learning in terms of a system (e.g. individuals with tools in a social context) and hence was considered more relevant to this research than other possible theoretical approaches. AT is also an approach designed to analyze purposeful action, one that places tools and purposes explicitly in the frame of analysis, and which allows a "critical incident analysis" approach to work easily. Identifying contradictions has been central to the analysis of user interaction for this research.

All 31 interactive and passive VR experiences and interviews, recorded on 50 hours of video, were transcribed following a two-stage analysis process: firstly, all the transcripts were annotated to highlight occurrences of critical incidents, breakdowns or contradictions; then, thematic categories were formed based on key incidents, i.e., problems with the largest number of breakdowns / contradictions. The following thematic categories were formed: **incidents caused by technical/usability problems** (approx. 5 incidents), **incidents caused by observer intervention** (approx. 9 incidents), **incidents caused by the problem of ordering fractions** (approx. 24 incidents), and **incidents caused by the problem of confusing the denominator with the result** (approx. 14 incidents). Excerpts were then extracted for more detailed consideration.

The examples presented in each of the thematic categories below are extracted from the IVR and PVR conditions on the basis of their being representative of the themes they illustrate. They were also selected because they correspond to participants that were more able to verbalise the reasons of their actions. Despite the fact that contradictions occurred on the same themes and learning problems in the non-VR condition, the activity with the LEGO bricks did not provide opportunities for resolution. In other words, system feedback in the form of constraints and guided discovery was not available to the participants, who had no way of knowing whether their solution for each task was right or wrong. Therefore, the participants' objective of designing a correct playground with LEGO was not reached, since all but one participant (the one being the only child who performed well on the pre- and post-tests) did not complete a correct playground in its entirety.

4.2.1.1 <u>The problem of ordering fractions</u>. Ordering fractions was difficult for many of the participants in the study, as expected. Both the extensive reference to this problem in the literature and the observations from the pilot study

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2017.2737983

had pointed to the fact that students tend to apply whole number strategies when comparing fractions. Furthermore, researchers [7] have found that this problem is amplified when the ordering task is embedded in a verbal problem-solving situation, as with the way the tasks were communicated to the participants in this study.

Hence, it is not surprising that most examples of conceptual conflict occurred with the swings task, which involved increasing the 3 x 4 area of the swings by comparing two fractions (the fractions 1/3 and 1/4) and choosing the number that represents the larger amount. Two examples from the transcripts were chosen to illustrate this theme, one from the IVR condition and one from the PVR condition. Each example illustrates a different kind of breakdown within the activity system.

Example 1- Jenny (IVR)

Jenny is a 9 year-old girl participating in the interactive VR condition (Fig. 1). Her ultimate goal was, as for all children in the study, to correct the playground. Her immediate goal, or action, in this example was to correct the area of the swings. The tools available for her to perform her action were all the possibilities provided by the interactive VE at the operation level (e.g. navigate freely in the VE with a joystick, pick and place blocks with the use of a button, switch between construction modes with a button, switch between ground level and top-down views, etc.). When required to choose between a third and a fourth of twelve, she incorrectly chose one fourth as the fraction that results in the larger number. She picked three blocks from the central "blocks pool" and attempted to fit these three blocks in the correct place so as to complete the task. Jenny tried out various solutions before realising, through an approach of reflection that was guided by her recalling previous experience of system feedback that she should have chosen one third instead of one fourth.

Jenny could make the calculation (i.e. she knew that a third of twelve gives four) but she was not able to respond to the ordering problem, so part of her rule set was correct but not all of it. System feedback led her to reach her goal by incorporating new rules to her existing set of rules, in other words, by integrating new pieces of knowledge to her existing model. Consequently, her response to Question A5a on her post-test (which was the exact same exercise of comparing one third and one fourth) was now correct. However, her response to Question A5b (comparison of two fractions with different numerators) remained incorrect. This indicates that despite her ability to complete the task in the virtual environment by ordering fractions of the same numerator, she was not able to transfer and generalise this rule to a similar problem on the post-test but with fractions with different numerators. She was able to resolve the contradiction but only partially. Therefore, it is not certain that the outcome of this process was achieved on a conceptual level, i.e. that deep understanding of concepts of order in fractions was gained.

Example 2 – Lisa (PVR)

The effect of system feedback on children's responses was observed in the passive VR condition as well, even though in the PVR case the children did not have a chance to respond kinaesthetically to such feedback and to try things out. They did, however, respond verbally, by predicting how many blocks the robot would be placing, thus "guiding" the robot in placing the blocks, and explaining the robot's actions and operations to the observer after the task. Lisa's response in the case of the swings was similar to Jenny's with the only difference that Jenny could act out her actions. In Lisa's case, the operations were performed by the robot and Lisa was restricted to observing the robot and telling the observer what the robot ought to be doing or why he had done what he had done. There is a division of labour here in the sense that the robot is taking on part of the work of fixing the swings area by performing the tasks in VR; Lisa, in her view, is aiding in this process by predicting what ought to be done. Hence, when the red bird finishes explaining the rule for the swings area, Lisa immediately "directs" the robot to "do one fourth because it's the bigger". The robot has finished placing all four blocks in the row and the swings are completed successfully. Lisa had made the same mistake as Jenny in picking one fourth of twelve as the fraction that gives more blocks and admitted that she had chosen it because of the number four in the denominator being a bigger number than three.

Later in the interview, the researcher asked Lisa to try to explain again how many blocks come out of one third of twelve and how many blocks come out of one fourth. This time she correctly responded that one third of twelve is four and one fourth of twelve is three and then recalled her mistake by realising that "oh yeah so one third is bigger". When asked why she thought that one fourth is bigger she laughed saying that "well, because four is bigger".

Lisa's initial rule set is incorrect as she is not able to respond correctly neither to the fractions calculation nor to the ordering of fractions. The visualisation of the area and her up-to-now experience of system feedback helped her to predict what the feedback messages from the system would be and thus decide where the robot would be placing the four blocks, i.e. on the only four tiles that were available. The robot's actions confirmed her choice and predictions and the playground was completed successfully with Lisa's initial conceptions remaining as they were before she entered the experience (as confirmed by her incorrect response to Question A5a on both the pre- and the post-test). However, when followed up by the observer, Lisa was able to explain the problem correctly and show that she understood why she had been mistaken. Lisa's contradiction with her rules is not resolved during the VR experience but only after she has finished with it, through prompting by the observer. So there is no contradiction between Lisa and the tool (in fact, in the PVR cases, there is no interaction between Subject and Tool at all) nor between Lisa and Division of Labour (the robot). This is a case where it is the observer's intervention that caused a revision in the participant's rules. The participant shows that she is not capable of resolving her misconceptions unaided. The observer, in this case, steps out of her role and by questioning Lisa, essentially supports her unintentionally, creating a Zone of Proximal Development [34]. Nevertheless, Lisa's revised understanding, has no long term effect on how she thinks about fractions since, in the post-test, she reverts to her previous misconceptions.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TETC.2017.2737983

4.2.1.2 Using the denominator as the answer. More than half of the participants in the study used the denominator of a fraction as the resulting number required by the task. This problem was faced with two of the playground elements, the slide which had to be increased by one fifth of ten, and the monkey bars, which had to be decreased by one sixth of twelve. The pre-test and the post-test both included questions - particularly Questions A2 and A8 - that required such calculations.

Example 1- Annie (IVR)

Annie, a 10 year-old participating in the IVR condition and attempting to correct the monkey bars area, believed that the answer to 1/6 of 12 was 6 but wasn't sure about it.

O.: Do you remember the calculation that you're supposed to make?
Annie: Uhm, yeah, one sixth of twelve
O.: How much do you think it is?
Annie: I thought it was six.
O.: And how many [*blocks*] are the monkey bars?
Annie: Six.
O.: So... if you take six away...
Annie: ... there'll be nothing
O.: What do you think the problem is? [...]
Annie: I'm just gonna try taking blocks away. [*She starts removing blocks and clicking after every action until they turn into monkey bars.*]
O.: So what was the correct answer then?
Annie: Two.
O.: Does that make sense to you?
Annie: Only cause of the two six. Two times six is twelve.

Annie has an initial conflict with her own rules R1 and with what she sees in the VE. Following consecutive trials for which she relied on system feedback, she was able to reach the answer, complete the task and then to actually explain the answer and resolve her contradiction.

This is an example of the interactive VR system aiding in problem-solving; Annie is able to solve the problem but does not show to have a theoretical understanding of the solution -there is no discourse in her process of trying to solve the problem that shows intentionality of her actions. Nevertheless, she is able to give a conceptual explanation in the end, which persisted to the post-test, since the result of her Question A2 was now correct. This is evidence of learning through problem-solving that was supported by the interactive VE.

Example 2 – David (IVR)

Like Annie, David participated in the IVR condition giving the same incorrect response to the monkey bars problem. Even though the problem is the same, the examples differ with respect to the resolution of the contradiction. David is a very talkative and animated boy who admits to playing computer games several hours per day, every day. He is a very competent player, confirmed by the computer log file which shows that he completed the activity in 23 minutes, even though he "didn't know his fractions" (he scored only 4 correct questions out of 11 on the pre-test).

David's is another example of solving the problem by trial and error supported by the interactive VE. His incorrect yet firm belief that 1/6 of twelve is found by subtracting 6 from 12 is not altered even when the observer essentially "warns" him that if he leaves no blocks he will not have monkey bars in the playground and thus risks not achieving his goal to complete a correct playground. He chooses to ignore the dominant rules of the community and goes on with executing his initial idea of taking away all the blocks. In this process, however, he tries things out until he discovers the solution, i.e., he keeps on clicking until he is able to leave the correct number of blocks on the

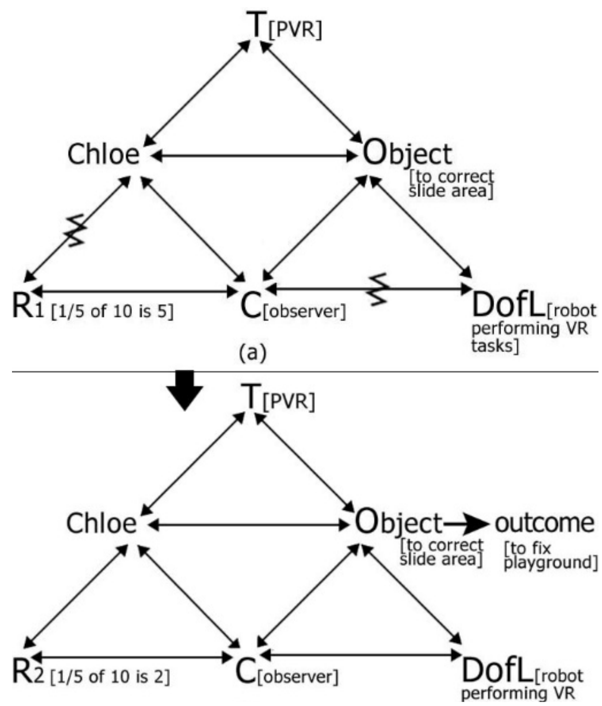

Fig. 4 An activity system illustrating in (a) the contradiction between Chloe's beliefs (a breakdown between Chloe and R1) and the beliefs of the community of which the robot is part. In (b) Chloe is able to reassess her model and resolve the contradictions, forming a new rule set R2 in which ten is divided by 5.

ground and complete the task.

His actions are performed with great skill and confidence in navigating to and from the pool and picking/placing blocks, a result of his experience in playing computer games, as he himself reported ("not because of fractions but because of computer games"). His confidence in using the tool suggests that he reverts to that method of solving a problem which he knows that he is good at rather than the conceptual process of trying to work out the fractions, which he admits that he is not good at. The tool (the interactive VE) was able to support his problem-solving activity but not his conceptualisation. Thus, he was able to get through the VR task and achieve his objective but there is no evidence of revision of his misconceptions. This is also reflected in his post-test scores, which remain low on Question A2 and on similar questions, indicating that his activity in the VE made no difference to his conceptual understanding of fractions.

Example 3 – Chloe (PVR)

Chloe, like David, was confident that increasing the slide area by one fifth of ten would mean adding five blocks. But when the robot completed the slide area correctly by placing only two blocks, a contradiction occurred between Spike's action and Chloe's model, or, more precisely, between Chloe's beliefs and the rules of the community of which the robot is part. The robot's actions took her by surprise, as up to that point she was certain about her rules. In AT terminology, the contradiction between her rules and the robot's rules is a breakdown between Chloe and the Division of Labour (represented by the robot) and is illustrated in part (a) of Fig. 4.

When thinking about what the robot did, Chloe reassessed her model and came up with a new one (in which the original number of blocks is divided by the denominator) that could later be generalised (Fig. 4). In fact, in the next task, which was to compare the two fractions (one third and one fourth) for increasing the area of the swings, she used her newly constructed model to come up with a correct response immediately. The form of her explanation of how the correct answer was derived indicated transfer of acquired knowledge to a new situation. Indeed, Chloe's activity system in this case presents a contradiction only in what Kuutti [17] calls the execution phase, which is quickly resolved when observing the robot's activity. Her reaching the object of the activity and eventually the outcome is evidence of conceptual change that is a result of her acquisition of the rules of the community through a division of labour. Furthermore, Chloe's performance in the post-test supports this evidence, as her post-test scores were improved over the pre-test scores on both problems of ordering fractions and dividing the correct numbers (i.e. without using the denominator as the response). The shifts in conscious attention from operations to actions allow for the reinterpretation of a situation in which objects become what Winograd et al. [35] refer to as "ready-at-hand".

4.2.2 Discussion of qualitative analysis results

The use of the analytical framework of Activity Theory provided the lens through which the critical incidents and internal contradictions - conflicts that required further attention were identified as possible indications of conceptual change. The importance of contradictions to Activity Theory is that they serve as "functions of a growing and expanding activity system", in other words, as indications of both discordance and, more positively, potential opportunities for intervention and improvement [2].

Indeed, what is evident from the examples presented here is that the representational cues of the VE, coupled with the interactive VR system's feedback mechanism, supported a certain type of activity and response which aided in problem-solving. The representational cues acted as visual forms of feedback for many participants, for example, judging if the area is a proper shape or guessing the number of blocks based on the available tiles and the surrounding space. Both cues and feedback created contradictions and then opportunities to predict contradictions. In this sense, the interactive VR environment was successful in supporting problem solving through a trial-and-error evolution strategy; consequently all participants in the IVR study were able to complete all the tasks.

The examples indicate that interactivity is well suited in facilitating the operations level, i.e. aiding the participant planning and problem solving. The question posed by this research, however, is whether the interactive properties of a VE, e.g. system feedback, can enable the transformation from conscious actions into operations, where planning and problem solving will have faded from the consciousness to give way to conceptual understanding.

As evidenced through the final example, the passive or guided VR condition provided the ability to reflect on the actions within the VE and try to make sense of the robot's activity. The robot acted as an additional level of mediation which, however, seemed in some cases to support the children's reflective thought, the ability to step back and consider a situation critically and analytically, with growing awareness of their own learning process. This finding agrees with the Vygotskian view that learning environments should involve guided interaction, permitting children to reflect on inconsistency and to change their conceptions [34]. Similarly, Smith et al. [33] argue that the shift from particular conceptions to more complex conceptual knowledge is a process relying on gradually linking ideas to other ideas in new ways. Davis [9] adds that giving students ample opportunities to reflect and identify where those links can be made (in the context of supportive learning environments) appears important in encouraging integrated conceptual understanding.

Vygotsky [34] also talked about the social context of learning through interaction with peers. Children often play or study in pairs or groups, helping each other in the process. The IVR and LEGO experiences were individual experiences in which the child was alone interacting only with the system and, unavoidably, with the observer of the study. In the case of the PVR, the robot assumed a similar role of an implicit probing entity embedded in the environment, albeit one with no conversational capabilities. Although no verbal exchange was or could be established between the child and the robot, the relationship between them (established through the child's careful observation of the robot's actions) took on a form of cooperation with a teacher or more able peer. In this case, the robot effectively acted as a mediator between the designer of the system and the learner, a kind of avatar, guiding and supporting the novice as he or she observed complex tasks, by emulating processes and behaviours typical of an expert. Hence, the robot is considered the "face" of the designer, acting as a more able peer through the suggestions and constraints that have been programmed into the system and which shape learner action. This support for reflection afforded by the passive VR condition was an unexpected outcome. The importance of reflection as a mechanism for conceptual learning is widely recognised by a number of scholars [8]. In the PVR case, the participant was required, both implicitly (by the robot through a division of labour) and explicitly (asked by the observer), to step back and try to make meaning out of the robot's actions. The instances that challenged participants' prior misconceptions to emerge were triggered by the contradictions or breakdowns between their beliefs and the beliefs of the community. In essence, this community, a social structure represented by the robot and the observer, showed to influence the participant's "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it" [11]. The examples collected of participants in the PVR condition provide evidence of this, whilst there were no comparable examples of conceptual change in the IVR condition. On the basis of this evidence, the PVR condition may support conceptual change which differs in that it is sustained (as in the example of Chloe).

5 Conclusion and Future Work

Given the gaps identified in earlier research on this topic, we have sought to provide a theoretically adequate and empirically grounded study of the role of interactivity in virtual environments for learning. We suggested a foundation for evaluating students' processes of learning within VEs by applying a descriptive method of interpretation that has not been used to our knowledge, in immersive VEs. In the process, fundamental methodological questions that go beyond this particular work were raised – not just the problem of working with children, but generally the problem of isolating subtle effects within an experimental framework. In the end, the combination of quantitative and qualitative methodology has proven indispensable and consequently opens up interesting directions for the study of deeper questions such as those concerning conceptual learning.

The results of the quantitative analysis could be interpreted as showing that the use of VR was beneficial. The qualitative analysis results indicate that activity based on the cues or feedback provided by the VE led participants to complete the tasks successfully in the IVR condition compared to the non-VR condition. Interactivity aided in promoting skill and problem solving and provided opportunities for contradictions to emerge. However, interactivity did not necessarily lead to resolution of these contradictions nor did it ensure that, if resolution was made, this was at the conceptual level. The qualitative analysis also revealed that a 'passive' VR form of experience (without interactivity, yet immersive), where the tasks were performed by a virtual robot observed throughout by the participant, afforded the potential to support resolution of contradictions in a way that encouraged reflection of the underlying learning problems. This guided form of interaction, rather than the fully interactive one, provided indications that learning takes place when shifting between experience–physical acts and reflection–mental acts.

From these findings, directions can be provided for future research work and guidelines on the design and development of user interfaces and interaction methodologies that cater to learners and support meaningful learning experiences. For example, detailed design paradigms for an educational VR environment could incorporate pedagogical guidelines that synthesise both a constructivist and a guided instructional approach. A practical design recommendation could involve harnessing the motivational power of interactivity to engage the participant in action, which would then be coupled with support for reflection. This dual prompting (prompts for action and for reflection) could be embedded in the same VE in a variety of forms, ranging from audio and visual feedback to the use of intelligent agents and storytelling mechanisms for vicarious action. This use of "layers of prompts" could be leveraged by other projects and could have an effect on future application design in which different levels are designed, similarly to the discursive, adaptive, interactive and reflective conversational framework proposed by [18].

Another, important, direction for further work being considered is to carry out longitudinal studies that examine the effect of interactivity in VR over the long term and with regards to other parameters that come into play, such as presence and body movement [32]. Learning is conceived to be a gradual process that cannot be studied in a brief experience, no matter how powerful that experience may be. Researchers studying the effect of informal interactive exhibits on learning [25] confirm that "even a highly attractive and stimulating show lasting 20 minutes won't lead to fundamental changes in a student's learning experience". Learning has to be embedded in a broader context and the interactive experience itself should be embedded in a much more extended learning process, where already acquired knowledge can be applied to new problems, exercised in new contexts, expanded or corrected, and students become motivated to seek further information. Additionally, despite an overall sample size of 50, the sample size per condition could be considered small. We plan to address this by conducting further studies with a larger sample using the VP on consumer VR headsets.

The key research question of this study has been whether interactivity, as an essential property of a VE, has an effect on conceptual learning. Studying the assumed relationship between interactivity and learning in VEs is not straightforward since a host of factors affecting learning come into play, including the learner's context and style. The analysis of the empirical studies suggests that interactivity promotes skill and problem solving and-most importantly with regards to the research question-that it can provide opportunities for contradictions to emerge.

On the other hand, a guided form of experience, as in the case of the VE where tasks were performed by a robot, showed the potential to support resolution of contradictions in a way that may encourage a reflective process. Here, an important parameter in the activity system is the social structure that forms a critical element for guidance and prompting. In the VP, this social structure was embodied in the robot that led the activity, thus encompassing an implicit instructional role. Additionally, the observer, also part of this social structure, had a positive role on some children's learning, even if unintentionally. What this suggests is that the social structure may be more important than interactivity, on its own, in supporting the process from problem solving to the making of meaning. Yet, in combination, interactivity and guided activity may be a powerful scaffolding tool to support reflection and sustained conceptual understanding.

With VR's recent resurgence in prominence and imminent mass consumption, the understanding of how humans interact in immersive VEs regains relevance, and can aid the broader community and practitioners in designing and engineering interactivity for training in formal or informal educational systems and contexts.

Acknowledgment

We extend our deep appreciation to Dr. Martin Oliver for his guidance. We also thank all adults and children who helped in the development and evaluation. The studies for this research were approved by the UCL Committee on the Ethics of Non-NHS Human Research, Study No. 0171/001.

References

[1] R. Bakeman and J. M. Gotman, *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge, UK: Cambridge University Press, 1997.

[2] S. A. Barab, M. A. Evans, and E. Baek, "Activity Theory as a Lens for Characterizing the Participatory Unit," in *Handbook of Research on Educational Communication and Technololgy*, D. H. Jonassen, Ed. Mahwah, NJ, USA: Lawrence Erlbaum, 2001, pp. 199–214.

[3] E. Bialystok and J. Codd, "Representing quantity beyond whole numbers: Some, none, and part," *Can. J. Exp. Psychol.*, vol. 54, no. 2, 2000.

[4] S. Bødker, "A human activity approach to user interfaces," *Human-Computer Interact.*, vol. 4, pp. 171–195, 1989.

[5] J. Cassell and H. Jenkins, "From Barbie to Mortal Kombat: Gender and Computer Games." The MIT Press, Cambridge, Massachusetts, p. 360, 1998.

[6] M. T. H. Chi, "Quantifying Qualitative Analyses of Verbal Data: A Practical Guide," vol. 6, no. 3, pp. 271–315, 1997.

[7] K. Cramer, M. J. Behr, T. R. Post, and R. Lesh, "Rational Number Project: Fraction Lessons for the Middle Grades," *Evaluation*, vol. 2, p. 190, 2003.

[8] M. W. Daudelin, "Learning from experience through reflection," *Organ. Dyn.*, vol. 24, no. 3, pp. 36–48, 1996.

[9] E. A. Davis, "Scaffolding Students' Reflection for Science Learning," University of California, Berkeley, 1998.

[10] C. J. Dede, M. C. Salzman, and B. R. Loftin, "MaxwellWorld: Learning Complex Scientific Concepts Via Immersion in VR," 1996, pp. 22–29.

[11] J. Dewey, *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Lexington, MA: Heath, 1933.

[12] Y. Engeström, *Learning by Expanding*. 1987.

[13] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports As Data*, Revised Ed. MIT Press, 1993.

[14] E. Hornecker and A. Dünser, "Of Pages and Paddles: Children's Expectations and Mistaken Interactions with Physical-Digital Tools," *Interact. Comput.*, vol. 21, no. 1–2, pp. 95–107, 2009.

[15] L.-E. Janlert and E. Stolterman, "The Meaning of Interactivity—Some Proposals for Definitions and Measures," *Human–Computer Interact.*, p. 07370024.2016.1226139, Sep. 2016.

[16] van Kesteran, E. H. Ilse, M. M. Bekker, A. P. O. Vermeeren, and P. a. Lloyd, "Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects," 2003.

[17] K. Kuutti, "Activity Theory as a potential framework for HCI research," in *Context and Consciousness: Activity Theory and Human Computer Interaction*, B. Nardi, Ed. Cambridge, MA: MIT Press, 1996, pp. 17–44.

[18] D. Laurillard, *Rethinking University Teaching: a conversational framework for the effective use of learning technologies*, 2nd ed. London: RoutledgeFalmer, 2002.

[19] P. Markopoulos and M. Bekker, "On the assessment of usability testing methods for children," *Interact. Comput.*, vol. 15, no. 2, pp. 227–243, Apr. 2003.

[20] T. Moher, A. E. Johnson, S. Ohlsson, and M. Gillingham, "Bridging Strategies for VR-Based Learning," 1999, no. 6, pp. 536–543.

[21] B. Nardi, Ed., *Context and Consciousness: Activity Theory and Human Computer Interaction*. Cambridge, Massachusetts: The MIT Press, 1996.

[22] J. Nielsen, "Getting access to what goes on in people's heads? - Reflections on the think-aloud technique," in *NordiCHI*, 2002, pp. 101–110.

[23] N. Otero, Y. Rogers, and. du Boulay, B., "Is Interactivity a Good Thing? Assessing its benefits for learning," *Praxis (Bern. 1994).*, pp. 1–5, 1995.

[24] N. Parés and R. Parés, "Interaction-Driven Virtual Reality Application Design," *Presence Teleoperators Virtual Environ.*, vol. 10, no. 2, pp. 236–245, 2001.

[25] U. Ritterfeld, R. Weber, S. Fernandes, and P. Vorderer, "Think Science! Entertainment Education in Interactive Theaters," *ACM Comput. Entertain.*, vol. 2, no. 1, p. Article 2, 2004.

[26] W. Robinett, "Interactivity and individual viewpoint in shared virtual worlds: the big screen vs. networked personal displays," *ACM SIGGRAPH Comput. Graph.*, vol. 28, no. 2, pp. 127–130, 1994.

[27] Y. Rogers, M. Scaife, S. Gabrielli, H. Smith, and E. Harris, "A Conceptual Framework for Mixed Reality Environments: Designing Novel Learning Activities for Young Children," *Presence Teleoperators Virtual Environ.*, vol. 11, no. 6, pp. 677–686, 2002.

[28] M. Roussos, A. E. Johnson, T. G. Moher, J. Leigh, C. A. Vasilakis, and C. R. Barnes, "Learning and Building Together in an Immersive Virtual World," *PRESENCE J.*, vol. 8, no. 3, pp. 247–263, 1999.

[29] M. Roussou, M. Oliver, and M. Slater, "The Virtual Playground: an Educational Virtual Reality Environment for Evaluating Interactivity and Conceptual Learning," *Virtual Real.*, vol. 10, no. 3–4, pp. 227–240, Oct. 2006.

[30] M. Roussou, M. Oliver, and M. Slater, "Exploring Activity Theory as a Tool for Evaluating Interactivity and Learning in Virtual Environments for Children," *Cogn. Technol. Work*, vol. 10, no. 2, pp. 141–153, 2008.

[31] J. W. Schofield and K. Anderson, "Combining quantitative and qualitative components of research on ethnic identity and intergroup relations," J. S. Phinney and M. J. Rotherarn, Eds. Newbury Park, CA: Sage, 1987, pp. 252–273.

[32] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 364, no. 1535, pp. 3549–3557, Dec. 2009.

[33] J. P. Smith III, A. A. DiSessa, and J. Roschelle, "Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition," *J. Learn. Sci.*, vol. 3, no. 2, pp. 115–163, Apr. 1994.

[34] L. S. Vygotsky, *Thought and Language*. Cambridge, MA: MIT Press, 1986.

[35] T. Winograd and C. Flores, *Understanding computers and cognition: A new foundation for design*. Reading, Massachusetts: Addison-Wesley, 1986.

[36] C. Youngblut, "Educational Uses of Virtual Reality Technology," Institute for Defense Analyses, Alexandria, VA, 1998.

[37] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. London, New York: Chapman and Hall, 1989.

Maria Roussou is an Assistant Professor in Interactive Systems at the National and Kapodistrian University of Athens. She holds a PhD (2006) in Computer Science from the University of London, a Master in Fine Arts (MFA -1997) in Electronic Visualization from the University of Illinois at Chicago, and a M.Sc.(1997) in CS also from UIC. She founded in 2003 makebelieve.gr. Previously, she established (in 1998) and directed (until 2003) the VR Department at the Foundation of the Hellenic World. She is Chair of the Greek ACM SIGCHI Professional Chapter and the recipient of the 2013 Tartessos Award for Virtual Archaeology.

Mel Slater is an ICREA Research Professor at the University of Barcelona and part-time Professor of Virtual Environments at UCL. He has been involved in research in VR since the early 1990s. He was awarded the IEEE VR Career Award in 2005, and recently completed a European Research Council Advanced Grant on VR. He is Field Editor of Frontiers in Robotics and AI. He has contributed to the scientific study and technical developments of VR and to the use of VR in other fields, notably psychology and the cognitive neuroscience of how the brain represents the body.