

# Tangibles in the Balance: a Discovery Learning Task with Physical or Graphical Materials

Paul Marshall<sup>1</sup>, Peter Cheng<sup>2</sup> and Rosemary Luckin<sup>3</sup>

<sup>1</sup>Department of Computing  
The Open University  
Milton Keynes, MK7 6AA  
p.marshall@open.ac.uk

<sup>2</sup>Department of Informatics  
The University of Sussex  
Brighton, BN1 9QH  
p.c.h.cheng@sussex.ac.uk

<sup>3</sup>London Knowledge Lab  
23-29 Emerald Street  
London, WC1N 3QS  
r.luckin@ioe.ac.uk

## ABSTRACT

An assumption behind much work on the use of tangibles for learning is that there are individual cognitive benefits related to the physical manipulation of materials. However, previous work that has shown learning benefits in using physical materials often hasn't adequately controlled for the covariates of physicality.

In this paper, we describe a study where we compared the effects on adults' discovery learning on a balance beam task of using either physical or graphical materials and with either control or no control over the design of experiments. No effects were found of either the type of learning material or the level of control over the experimental design.

## Author Keywords

Tangibles, physical, graphical, balance beam task

## ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):

## INTRODUCTION

Much work on tangible interfaces has suggested that they might hold promise for supporting learning (e.g. [20, 22]). A first exploratory phase of tangibles research primarily involved the design of a number of tangible systems with novel physical interaction mechanisms. However, there was initially little evidence to support assumptions about the learning benefits of working with tangible interfaces [11].

Current work has been more grounded in theory and empirical exploration and has advanced our understanding of learning through tangible interaction on a number of

fronts. For example, Antle and co-workers [1] have started to experiment with priming conceptual metaphors through physical movements to facilitate children's understanding of abstract concepts. Horn's [7, 8] research on physical programming has described some of the advantages of using low-tech robust physical artefacts in formal and informal learning scenarios. A number of researchers have begun to detail the representational properties of tangibles and physical artefacts (e.g. [4, 14]); and video-analytic studies have described some of the influences of using physical materials on the organisation of social interaction (e.g., [5, 13]).

The study described in this paper aimed to address one of the more fundamental underlying assumptions about tangible interfaces and learning: whether and why the manipulation of physical artefacts might have individual cognitive influences on learning.

## Physical interaction and learning

There is an assumption within the literature on tangibles for learning and more generally within writings on education that conceptual learning benefits accrue with the manipulation of physical artefacts [11]. For example, the 'hands-on' learning approach (e.g., [6]) emphasises the role of active experimentation with physical materials, often in the context of science education. Claims have also been made about the utility of concrete manipulative materials (e.g., [19]). Both formal systems of manipulatives, such as Dienes Blocks and Cuisenaire Rods, and informal types such as paper clips, chocolate bars and coins are used in education. However, the few reviews of research on the use of manipulatives and "hands-on" science have found inconsistent effects ([15, 18]).

Triona and Klahr [21] have argued that much of this work has had a broad educational philosophy in mind and therefore hasn't adequately controlled for covariates with physicality. One possibility is that educational approaches with physical materials tend to give learners more control over the educational experience; learners are the ones using the materials and are therefore likely to have more influence over what they do with them. This could both increase the learner's level of motivation and potentially provide them with information more relevant to their

current understanding of the domain, and thus be more conducive to constructivist learning.

To investigate this issue we designed a study, described below, to separate the effects of control and manipulation and to empirically test the benefits of manipulating physical materials compared to manipulating a graphical representation with a mouse for a learning task where we anticipated physical materials to have an positive effect. Adult participants carried out experiments to discover the rules which guide the behaviour of a balance beam.

To increase the likelihood of finding potentially subtle effects of using physical materials, sensitive measures of learning were used and a version of the balance beam task with continuous quantities was used to discourage participants from using mathematical strategies to abstract away from the physical relationships.

## METHOD

### Balance Beam Task

The balance beam task is a standard within developmental and cognitive psychology (e.g., [9, 17]). It was chosen for two reasons. Firstly, it involves reasoning about physical variables and is therefore in a domain where information obtained through physical manipulation might be expected to facilitate performance (cf. [11]). Secondly, the performance of participants on the balance beam task can be described as the application of one of a series of increasingly sophisticated rules: Rule I, Rule II, Rule III, Qualitative Proportionality, Compensation Rule or Rule IV [17]. For example, Rule I involves always predicting that the side with the largest weight will go down, irrespective of the distance from the fulcrum; Rule IV (the correct rule) involves calculating the product of the weight and its distance from the fulcrum, comparing the product on each side of the beam and predicting that the side with the largest will tip down, or if they are equal predict that the beam will balance. These rules allow the use of more sensitive measures of learning than all-or-none pre and post-test scores. As well as comparing total scores, the pattern of each participant's responses to pre and post-test questions can be analyzed to deduce the most likely rule that they are using to answer the questions. The rule used provides an ordinal score of the sophistication of a participants' understanding of the concept of balance on a scale ranging from 0 (random guessing) to 6 (Rule IV). Participants' test performance was assigned to a particular rule using maximum likelihood scores (see [12] for details).

The conventional version of the balance beam task uses discrete measures of weight and distance (e.g. weights of uniform size positioned on pins at unit distances from the fulcrum). A previous study using discrete quantities [12] found learning to have occurred after using the balance beam apparatus, but no benefit in using physical compared to graphical materials. The aim of the experiment reported here was to make manifest any potential cognitive effects

of physicality and control on conceptual learning by reducing the likelihood of participants using mathematics to reason about the behaviour of the beam. In a study carried out by Schwartz, et al. [16], ten year old children perform at the level of five-year-olds in a version of the balance beam task with continuous quantities of weight and distance. They argue that the continuous variables are difficult to quantify and that quantification is a prerequisite of using mathematical strategies. It was reasoned that discouraging adult participants from quantifying in this way might focus their attention on developing a more implicit understanding of the behaviour of the beam grounded in physical interaction.

### Participants

Thirty-four subjects responded to an advert and were paid for participation in this experiment. Thirty-two were female and two were male. The median age was 20 (range 18-46). Twenty-nine of the participants were students. None had taken part in a course at a higher level than GCSE (completed at age 16 in the UK) or equivalent in mathematics, physics, engineering, or any other of which mechanics was a component.

### Design

This experiment employed a pre-test—intervention—post-test design. There were 2 (materials: physical or screen) x 2 (with-control or no-control) experimental conditions at the intervention phase. Participants in the physical condition used a physical balance beam, while those in the screen condition worked with a graphical balance beam controlled with a mouse. Participants in the with-control condition were asked to devise and carry out experiments to work out the rule controlling the behaviour of the beam. Those in the no-control condition were given experiments to carry out.

### Materials

Both a physical and an on-screen balance beam were constructed for use in this experiment. The physical beam comprised a 1 metre long wooden arm to which a wood and lead counterweight was attached perpendicularly at the midpoint. The arm was attached to a wooden base by a wooden axle and able to swing freely. A strip of rubber tape was applied to the top of the arm to provide a high friction surface for the weights to hang from.

Two transparent plastic tubs were used as weights for the balance beam apparatus. The tubs had plastic handles, which could be hung in any position along the length of the balance arm. The tubs could be filled with sand from a large container. The counterweight was sufficiently heavy that the beam remained in balance when an empty tub was positioned at the end of it.

The screen beam was constructed in Macromedia Director using photographs of the physical apparatus so as to be visually similar. A screenshot from the screen apparatus is shown in figure 1. An 'add sand' and 'remove sand' button appeared underneath each side of the beam, which the



**Figure 1: Screen balance beam with continuous quantities**

participants used to change the quantity of sand in the tubs. Holding down a button caused the level of sand to slowly increase or decrease. The tubs could be moved along the beam by clicking and dragging them. They could not be moved beyond the end of the beam or past the midpoint.

The beam rotated at a speed proportional to the level of sand and position of the pots on the beam. Thus, if a nearly empty pot was positioned near the middle of the beam on the left hand side, and a full pot was positioned at the end of the beam on the right hand side, the right hand side of the beam would move down rapidly.

Every time a change was made to the balance beam apparatus, a line was appended to a log file giving the amount of sand in, and distance from the fulcrum of both of the pots. This was used in the analysis.

### Tests

To provide a sensitive measure of participants' level of understanding and to increase the likelihood of finding a physicality effect, three different tests were designed: a test of motor performance, a written test with continuous quantities. It was reasoned that participants would be most likely to use explicit mathematical strategies on the test with discrete quantities and that the test with continuous quantities would tap into a level of comprehension between motor performance and discrete representations (cf. [16]).

#### *Motor performance test*

The motor performance test was designed to measure how accurately participants could actively position different weights to balance out the beam without feedback. There were two variations of the motor performance test, one for participants in the physical conditions and one for participants in the screen conditions so that participants' learning was tested with the same apparatus used in the experiment phase. In both variations, the experimenter filled two of the pots with varying quantities of sand and positioned one at a point on the beam. The amount of sand in each of the pots was one of either  $\frac{1}{4}$  of a pot,  $\frac{1}{3}$  of a pot,  $\frac{1}{2}$  a pot,  $\frac{2}{3}$  of a pot,  $\frac{3}{4}$  of a pot, or a full pot. The pots were positioned either  $\frac{1}{4}$ ,  $\frac{1}{3}$ ,  $\frac{1}{2}$ ,  $\frac{2}{3}$ ,  $\frac{3}{4}$  or all of the distance from the fulcrum to the end of the beam. The task for the



**Figure 2: A screenshot from the physical motor performance test**

participant was to position the other pot where it would balance out the beam. These values were not represented numerically to the participants and therefore they had to judge how full each pot was and at what the proportion of the distance from the fulcrum the first pot had been placed.

After positioning the first pot, the experimenter held the balance beam horizontally and the participant was asked to touch the second pot to the beam where they thought it would balance out the beam, but without putting the full weight of the pot down (as the motor performance test was used in the pre-test and post-test phases it was important that the participants received no feedback as to whether they were correct that may have influenced their other responses). Figure 2 shows one of the participants positioning a pot on the beam.

For the physical version, participants were videoed positioning the weights on the beam and screen-captures were taken at the point when they touched the pot to the beam. For each of these images the distance between the fulcrum and the pot and the fulcrum and the end of the beam were measured. The distance of the pot from the fulcrum was calculated as a percentage of the length of the beam.

For the screen version of the test, an altered version of the Director movie was used where the beam did not move when changes were made. The experimenter set the levels of sand in both of the pots and positioned one of them. The participant was then asked to position the other pot where they thought it would balance out the beam. As the beam did not move, they received no feedback on whether they were correct. The distance of the pot from the fulcrum was calculated as a percentage of the length of the beam.

Accuracy for the motor performance tests was determined by calculating the difference between the percentage distance from the fulcrum where the pot was positioned and the correct percentage distance. Thus, if the correct distance was 25% and the pot was positioned at 30.5%, then the score would be 5.5 for that item. The more accurate participants were in the motor performance tests, the smaller their score.

### Written tests

Two sets of written tests were constructed which used either continuous or discrete quantities. The tests with continuous quantities were designed to investigate participants' learning about the concept of balance at a level between that measured by the sensory-motor tests and the written tests with discrete quantities, i.e., separable from the context in which it was learned but still using continuous quantities and relying on visual information rather than both visual and haptic information (as measured by the motor performance test). The tests with discrete quantities were designed to investigate participants' abstracted mathematical understanding of the behaviour of a balance beam.

Four written tests of ten items were constructed using continuous representations of sand and distance and four with discrete representations. Each item comprised a diagram showing a balance beam with two pots filled with varying quantities of sand. Pots were always  $\frac{1}{4}$ ,  $\frac{1}{3}$ ,  $\frac{1}{2}$ ,  $\frac{2}{3}$ ,  $\frac{3}{4}$  or completely full and positioned at similar proportions of the distance from the fulcrum to the end of the beam. Participants were asked to predict whether the beam would tip left, remain in balance or tip right.

For both types of written test, total scores were calculated and compared between groups. Likelihood scores were calculated for each of the balance beam rules (random answering, Rule I, Rule II, Rule III, QP, Compensation Rule or Rule IV) and the most likely rule was selected to describe participants' performance at both pre- and post-test. Rule use was described on an ordinal scale of sophistication from 0 (random answering) to 6 (Rule IV) and compared across groups.

### Procedure

Participants were assigned to either the physical or screen conditions and either the with-control or no-control conditions.

Participants first worked with the experimenter to answer questions on one of the four motor performance tests. Those in the physical conditions used the physical apparatus, and those in the screen conditions used the modified Director movie. This stage of the experiment was video recorded for later analysis. They then completed one of the written tests with continuous quantities and finally one of the written tests with discrete quantities. The test

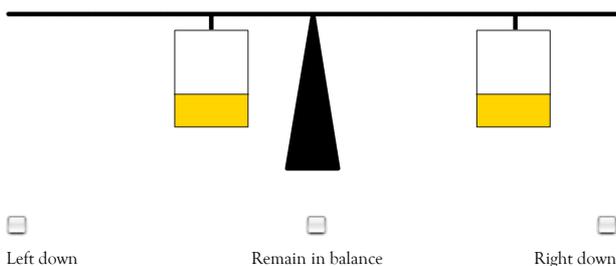


Figure 3: Item from written test with continuous quantities

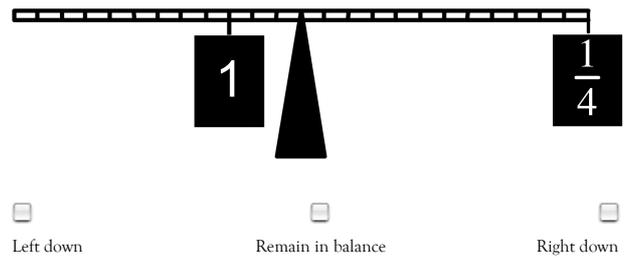


Figure 4: Item from written test with discrete quantities

version was counterbalanced across conditions for all three tests.

In the second stage of the experiment, participants in the physical conditions used the physical apparatus and those in the screen conditions used the screen balance beam. To allow investigation of the effect of learners having or not having control over their own discovery learning, there were differences in the procedure in the with-control and no-control conditions: with-control participants were asked to design and carry out their own balance beam experiments, while no-control participants were provided with experiments to carry out.

Participants in the with-control conditions were asked to experiment with the balance beam apparatus to try and work out the rule that determines whether the beam will balance, tip left, or tip right for all quantities of sand and distances. They were given six experiments to do so. An experiment was defined as changing the quantity of sand and/or the position of one or both of the pots. Participants were asked to make one definite movement with the each of the pots when positioning them rather than gradually sliding them along until they found the balance point; in the physical condition this meant picking up the pot and putting it down again on the beam. In the screen condition, this meant clicking on the pot and releasing it at the chosen location. Once they had completed an experiment, if the beam was almost in balance, they were then allowed to change the position or sand levels slightly in order to completely balance the beam.

Participants in the no-control condition also carried out six experiments, but rather than devising their own, they carried out one of four sets set of model experiments by copying the distances and quantities of sand represented in pictures. The four sets were counterbalanced across conditions. The experiments were completed in order of increasing complexity (cf. [17]) and together provided enough potential information for the participant to correctly deduce that the behaviour of the beam is predicted by balance beam rule IV. Again, if the beam was almost in balance, the participants were able to make small adjustments to completely balance the beam.

After completing the six experiments, participants completed another motor performance test, written test with continuous quantities and written test with discrete quantities. In total, the experiment took approximately forty-five minutes per participant.

## RESULTS

### Motor Performance tests

Mean motor performance pre and post-test scores are presented by group in table 1. As this score was the percentage distance from the correct position, the smaller the score, the more accurate the performance.

Material	Level of control	Mean score (standard deviation)	
		Pre-test	Post-test
screen	with-control	10.1 (6.3)	8.4 (3.1)
	no control	15.0 (8.5)	10.2 (6.7)
physical	with-control	12.6 (5.9)	12.3 (5.6)
	no control	11.7 (4.4)	9.5 (3.5)

**Table 1: Mean motor performance pre and post-test scores by condition**

Mean motor performance test scores were found to differ significantly from a normal distribution for the with-control group's pre-test scores,  $D(18) = 0.26$ ,  $p < .01$ , and for both with-control,  $D(18) = 0.27$ ,  $p < .01$ , and without-control,  $D(16) = 0.22$ ,  $p < .05$ , groups at post-test. A log correction was applied to the data before analysis.

A mixed ANOVA was carried out on corrected motor performance test scores with time of test (pre-test and post-test) as a repeated measure and material and level of control as between-subjects factors. A main effect of test time approached significance,  $F(1, 30) = 2.66$ ,  $p = .07$ ,  $r = 0.29$ . However, there was no significant interaction between test time and materials used,  $F(1, 30) = 0.45$ ,  $p > .05$ ,  $r = 0.12$ , test time and level of control over the design of experiments,  $F(1, 30) = 0.87$ ,  $p > .05$ ,  $r = 0.17$  or between test time, material and control,  $F(1, 30) = 0.04$ ,  $p > .05$ ,  $r = 0.04$ . This suggests that experimenting with the balance beam led to an improvement in motor performance test scores that approached significance regardless of whether screen or physical materials were used or whether participants had control over the design of experiments.

There was no significant main effect of material  $F(1, 30) = 1.22$ ,  $p > .05$ ,  $r = 0.20$ , suggesting that there was no general benefit in using either physical or digital materials on accuracy in the motor performance test.

### Written tests with continuous quantities

#### Test scores

Mean pre and post-test scores for written tests with continuous quantities are presented in table 2. A mixed ANOVA was performed on the scores for the written tests with continuous quantities with time of test (pre and post) as a repeated measure and material and level of control over experimental design as between-subjects factors. There was no main effect of time of test,  $F(1, 30) = 0.54$ ,  $p > .05$ ,  $r = 0.13$ . Nor was there an interaction of time of test

with material,  $F(1, 30) = 1.877$ ,  $p > .05$ ,  $r = 0.24$ , or with level of control over experimental design,  $F(1, 30) = 0.54$ ,  $p > .05$ ,  $r = 0.13$ . Therefore, participants' scores did not change between pre and post-test irrespective of what materials they experimented with or whether they had control over the design of the experiments.

Material	Level of control	Mean score (standard deviation)	
		Pre -test	Post-test
screen	with-control	7.4 (1.7)	6.9 (1.6)
	no control	7.4 (1.9)	7.5 (1.7)
physical	with-control	6.3 (1.7)	6.8 (1.0)
	no control	6.9 (1.4)	7.6 (1.8)

**Table 2: Mean scores out of a maximum of 10 for written test with continuous quantities by condition**

#### Rule use

Maximum likelihood scores were calculated for each of the balance beam rules outlined above for pre and post-tests for each participant. As rule used was an ordinal measure, the median was used as a measure of central tendency. The median rules used by participants in each condition are tabulated in table 3.

Material	Level of control	Median rule used (range)	
		Pre-test	Post-test
screen	with-control	5 (0 – 6)	4 (2 – 6)
	no control	5 (1 – 6)	4.5 (2 – 6)
physical	with-control	4.5 (0 – 6)	4 (2 – 6)
	no control	4 (2 – 6)	4.5 (3 – 6)

**Table 3: Median rules used for written tests with continuous quantities by condition**

A Wilcoxon signed-rank test found no overall differences in rule use between pre (Mdn = 4) and post-test (Mdn = 4),  $T = 137.5$ ,  $p > .05$ ,  $r = -0.04$ .

Condition	Median difference	Minimum	Maximum
Screen-control	0	-4	6
Physical-control	0	-4	3
Screen-no-control	0.5	-2	2
Physical-no-control	1	-3	3

**Table 4: Median change in level of rules used between written pre and post-tests with continuous quantities**

As a factorial comparison is not possible with non-parametric data, material and control conditions were combined to produce four experimental groups: screen-

control, screen-no-control, physical-control and physical-no-control. Median differences between rules used in the pre and post-tests are presented by condition in table 4. A comparison of these difference scores between groups with a Kruskal-Wallis one-way analysis of variance by ranks revealed no significant differences,  $H(3) = 0.89, p > .05$ .

### Written tests with discrete quantities

#### Test scores

Mean pre and post-test scores for written tests with discrete quantities are presented in table 5.

Material	Level of control	Test score (standard deviation)	
		Pre-test	Post-test
screen	with-control	7.8 (1.6)	8.3 (1.8)
	no control	7.5 (2.4)	8.0 (1.5)
physical	with-control	7.0 (1.9)	8.1 (1.6)
	no control	7.6 (0.7)	7.9 (1.2)

**Table 5: Mean scores for written test with discrete quantities by condition**

A mixed ANOVA was carried out on the pre and post-test scores with time of test as a repeated measure and with materials used and level of control over the design of experiments as between-subjects factors. There was a main effect of time of test: test scores were found to increase significantly between pre-test and post-test,  $F(1, 30) = 4.63, p < .05, r = 0.37$ . However, there was no interaction between either time of test and materials used,  $F(1, 30) = 0.11, p > .05, r = 0.06$  or time of test and level of control over the design of experiments,  $F(1, 30) = 0.31, p > .05, r = 0.10$ . Therefore, performance on the written test with discrete quantities improved for all groups, regardless of whether physical or on-screen materials were used or whether the participants had control over the design of experiments.

#### Rule use

Likelihood scores were calculated for each balance beam rule for every participant's pre and post-tests and the most likely rule was selected to describe performance on that test. The median rules used by participants in each condition are tabulated in table 6.

Rules used on the pre and post-tests were compared using a Wilcoxon signed ranks test. No significant difference was found between pre (Mdn = 5) and post-test (Mdn = 5),  $T = 92, p > .05, r = -0.14$ , indicating that in general, the sophistication of participants' rule use did not increase between the pre-test and post-test.

As with the tests with continuous quantities, differences in the level of balance beam rule used between pre and post-test were compared with a Kruskal-Wallis one-way analysis

of variance by ranks. This revealed no significant differences between conditions  $H(3) = 0.36, p > .05$ .

Material	Level of control	Median rule used (range)	
		Pre-test	Post-test
screen	with-control	5.5 (0 – 6)	6 (0 – 6)
	no control	5 (1 – 6)	6 (2 – 6)
physical	with-control	4.5 (3 – 6)	5 (3 – 6)
	no control	4.5 (2 – 6)	5.5 (3 – 6)

**Table 6: Median rules used for written pre and post-tests with discrete quantities by condition**

Control	Median	Minimum	Maximum
Screen—with-control	0	-4	6
Screen—no-control	0	0	2
Physical—with-control	0	-3	2
Physical—no-control	0.5	-3	3

**Table 7: Median differences in level of balance beam rule used between pre and post-test**

There was no change in the level of balance beam rules used by participants while answering questions on the written tests with discrete quantities after experimenting with the balance beam apparatus. Nor were there any differences evident between experimental groups. It therefore appears that participants scored higher in the post-test, but without using more sophisticated balance beam rules. Two possible explanations for this finding are firstly, that the less powerful non-parametric statistics simply failed to detect a difference in the level of rule used, or secondly, that participants applied rules more consistently in the post-test than in the pre-test.

This second possibility was tested by comparing the log-likelihood scores calculated for the most likely rule used in both the pre and post-test. The log-likelihood of the most likely rule used in the post-test was found to be significantly closer to zero ( $M = -1.45, SE = 0.22$ ) than that used in the pre-test (mean =  $-2.05, SE = 0.21$ ),  $t(34) = -2.19, p < .05, r = 0.35$ , and therefore participants applied rules more consistently in the post-test than in the pre-test. A possible explanation is therefore that participants were more likely to settle for one strategy in the post-test, where they may have switched between different strategies at pre-test. This greater consistency would appear to account for the increase in performance. If participants were using rules inconsistently in the pre-test, then the comparison between pre and post-test rules would not have been a valid one, and this may account for the negative finding.

## DISCUSSION AND CONCLUSIONS

The aim of the experiment described in this paper was to maximize the likelihood of finding an effect of manipulating physical materials on an individual discovery learning task by trying to encourage participants to focus on the physicality of materials and by using a suite of sensitive measures of learning.

Working with the balance beam led to a measurable improvement in participants' scores on the test with discrete quantities, but not in the sophistication of the rule used to answer the questions. The motor performance tests showed an improvement in accuracy between pre and post-test that approached significance. However, there was no improvement in participants' performance on the written test with continuous quantities.

That the participants' overall performance improved in the discrete test, but that the sophistication of the balance beam rules used in the same test did not improve would seem to suggest that participants learned to apply rules more consistently between pre-test and post-test. To test this hypothesis, the log-likelihood scores calculated when assigning participants' discrete test answers to the best-fit rule category were compared at pre and post-test. The log-likelihood scores at post-test were found to be significantly closer to 0 than those at pre-test, providing some support for this hypothesis

This raises the question, if the discrete test scores show that participants' became more consistent in their application of balance beam rules between the pre and post-test for the discrete test, why were they unable to transfer this increased consistency to the continuous tests? A possible explanation is that they tried to apply the same mathematical rules in these tests, but that they incorrectly quantified the distance or weight dimensions. Applying the rule to incorrect values would therefore lead to an incorrect conclusion.

Informal observations of participants' spontaneous dialogue while they were working on the six experiments in the intervention phase of the study and during the motor performance tests lends some support to this hypothesis: participants often incorrectly described quantities. In particular, they often misidentified  $\frac{1}{4}$  distances and weights as  $\frac{1}{3}$ s (and vice-versa) and  $\frac{3}{4}$  with  $\frac{2}{3}$ . However, as participants were not asked to talk aloud while carrying out the experiments and often didn't, a more systematic comparison was not possible. If correct, this explanation would complement existing work on the balance beam task, which has suggested a value in attempting to apply mathematical strategies related to the familiar possible structures, such as an additive or multiplicative relationships, provided by thinking in mathematical terms [16]. Improved understanding of a physical phenomenon would be predicted by the application of a mathematical strategy [16] Unlike the 10 and 11-year-old children by Schwartz and coworkers [16] however, it would seem that adults were able to attempt a mathematical solution to the

balance beam task even when they were discouraged from doing so by working with hard-to-quantify continuous values. In the absence of more systematic analysis to confirm it, this view remains speculative.

The main finding of this study is that although trying to maximize the likelihood of finding a physicality effect, no differences were found for any of the learning measures between participants using the physical and graphical versions of the balance beam apparatus. This provides evidence that the physicality of the materials in the balance beam task may have little effect on learning for adult participants. While it is not possible to say conclusively that there is no effect of physicality based on this study, given the relatively low number of participants and the high standard deviations in test scores, it provides little support for the notion that the manipulation of physical materials has cognitive benefits for learning, and is in line with related findings by Triona and Klahr [21].

Whether younger participants, with less mathematical facility, would derive cognitive benefits from using physical materials is an open question. Druyan [3] asked young children to predict the direction of movement of a balance beam and then gave them feedback either kinaesthetically, by asking the children to pick up the beam via a loop at its fulcrum, visually, or socially. She found the kinaesthetic feedback to be most effective for preschoolers, which seems to suggest that physical materials would be of use with this age group. However, it is also possible that the benefit over visual materials was simply due an increase in motivation or attention by playing an active role in the study.

No differences were found for any of the measures between the with-control and no-control groups. Thus, it would appear that this factor had little effect on participants' learning in this version of the balance beam task; participants were able to improve their performance on the discrete test regardless of whether they designed their own experiments or had the experiments given to them. Two possible explanations for this finding are proposed. Firstly, participants did not choose to conduct experiments that provided useful information about the behaviour of the balance beam. For example, they may have shown confirmation bias in designing experiments to support rather than disconfirm hypotheses. It is not possible to determine what kinds of strategy participants were using in designing experiments. A second possible explanation is that using mathematical strategies simplified the problem of inducing a balance beam rule such that participants were able to choose a solution with only limited feedback. Under this explanation, both conditions would have provided sufficient feedback for participants to induce a rule. However, it is not possible to choose between these alternatives here and is suggested as a focus for future work.

There has been an assumption within some work on tangible interfaces that cognitive benefits accrue from the

manipulation of physical artefacts, which are then augmented through digital augmentation (cf. [11]). The primary aim of the study described in this paper was to search for evidence of individual cognitive benefits in manipulating tangible materials when learning about the physical phenomenon of balance. No such learning benefits were found in this case.

This certainly does not mean that no such cognitive benefits will be found either in other learning domains, with alternative learning activities or with different groups of participants (e.g., young children). However, it does suggest that they should be demonstrated rather than assumed. This finding therefore contributes to an emerging body of research that is starting to pin down the social, cognitive and affective outcomes of learning through tangible interaction (e.g., [1, 4, 5, 7, 8, 13, 14]).

#### ACKNOWLEDGMENTS

We thank those who volunteered to participate in the study. We are grateful to our reviewers for helpful comments and to Yvonne Rogers and Eva Hornecker who read drafts of the paper.

This research was supported by an EPSRC Equator Project studentship held by the first author.

#### REFERENCES

1. Antle, A.N., Corness, G., & Droumeva, M., What the Body Knows: Exploring the Benefits of Embodied Metaphors in Hybrid Physical Digital Environments, *Interacting with Computers*, 21 (2009), 66-75.
2. Clark, A. Beyond the flesh: some lessons from a mole cricket. *Artificial Life*, 11, (2005), 233-244.
3. Druyan, S. A comparison of four types of cognitive conflict and their effect on cognitive development. *International Journal of Behavioral Development*, 25, 3, (2001) 226-236.
4. Edge, D. and Blackwell, A.F. Correlates of the cognitive dimensions for tangible user interface. *Journal of Visual Languages and Computing*, 17, 4 (2006), 366-394.
5. Fernaeus, Y., & Tholander, J. (2006). Finding design qualities in a tangible programming space. In *Proc. of CHI 2006*, pp. 447-456.
6. Haury, D. L., & Rillero, P. (1994). *Perspectives of Hands-On Science Teaching*. Columbus, OH: ERIC Clearinghouse
7. Horn, M. S. and Jacob, R. J. Designing tangible programming languages for classroom use. In *Proc. TEI 2007*, 159-162
8. Horn, M. S., Solovey, E. T., Crouser, R. J., and Jacob, R. J. Comparing the use of tangible and graphical programming languages for informal science education. In *Proc. CHI '09*, 975-984.
9. Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: an essay on the construction of formal operational structures*. London: Routledge
10. Klemmer, S. R., Hartmann, B., & Takayama, L. (2006). How bodies matter: five themes for interaction design. In *Proc. DIS 2006*, 140-149.
11. Marshall, P. Do tangible interfaces enhance learning? In *Proc. TEI 2007*, 163-170.
12. Marshall, P. *Physicality and Learning - Searching for the Effects of Tangibility in Scientific Domains*. Unpublished DPhil Thesis, University of Sussex
13. Marshall, P., Fleck, R., Harris, A., Rick, J., Hornecker, E., Rogers, Y., Yuill, N., and Dalton, N. S. Fighting for control: children's embodied interactions when using physical and digital representations. In *Proc. CHI 2009*, 2149-2152
14. Price, S. A representation approach to conceptualizing tangible learning environments. In *Proc. TEI 2008*, 151-158
15. Raphael, D., & Wahlstrom, M. The influence of instructional aids on mathematics achievement. *Journal for Research in Mathematics Education*, 20, (1989), 173-190.
16. Schwartz, D. L., Martin, T., & Pfaffman, J. How mathematics propels the development of physical knowledge. *Journal of Cognition and Development*, 6, 1 (2005), 65-88.
17. Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 46(2), 1-74.
18. Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, 20(5), 498-505.
19. Suydam, M. N. (1986). Manipulative materials and achievement. *Arithmetic Teacher*, 33(6), 10-32.
20. Terrenghi, L., Kranz, M., Holleis, P., & Schmidt, A. (2006). A cube to learn: a tangible user interface for the design of a learning appliance. *Personal and Ubiquitous Computing*, 10, 2 (2006), 153-158.
21. Triona, L. M., & Klahr, D. Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, 21, 2 (2003), 149-173.
22. Zuckerman, O., Arida, S., & Resnick, M. (2005). Extending tangible interfaces for education: digital montessori-inspired manipulatives. In *Proc. CHI 2007*, 859-868.

