

A comparison of the impact of 3 forms of “hands-on” activities for learners with different scientific reasoning abilities

Susan M. Fischer,¹ Elise Agra,² and Sian L. Beilock²

¹*Department of Physics, DePaul University, 2219 N. Kenmore Ave., Chicago, IL, 60614*

²*Department of Psychology, University of Chicago, 5848 S. University Ave., Chicago, IL, 60637*

We have developed a set of activities designed to aid student understanding of torque and center of gravity. The activities exist in three forms, each appropriate for a group member engaged in a “hands-on” classroom experience: embodied, involving balancing objects on one’s hands, traditional hands-on, involving a meter stick balanced on a fulcrum, and observation of an experimenter performing the embodied activity. Our study is outside of the classroom and includes two populations of learners with significantly different abilities at pre-test. We find that the different types of training impact the two populations differently. In particular, participants with lower overall pre-test accuracies who observe the embodied activity show poorer gains on torque questions requiring proportional reasoning than participants with hands-on or embodied training, while observer participants with higher overall pre-test accuracies achieve high gains on even the more challenging transfer problems.

Keywords: scientific reasoning, proportional reasoning, embodied cognition, torque, center of gravity

I. INTRODUCTION

The balance beam task, which consists of a stick or board resting on a fulcrum with objects of various weights hanging on either side, has been employed in a variety of contexts in cognitive science and science education research. For example, elementary school children whose gestures contradicted their verbal explanations of the balance beam system were more amenable to instruction than children whose gestures matched their explanations [1]. A balance beam is included as one of the scenarios used to assess proportional reasoning ability on the Lawson’s Classroom Test of Scientific Reasoning [2], and was included in a study that compared teenagers’ ability to solve proportional reasoning problems before and after training using either verbal instruction or physical manipulatives [3]. The task has also been used to compare adult participants’ ability to develop and apply the rules of torque and rotational equilibrium before and after intervention with a physical system versus a computer simulation [4].

A recent study by our research group explored the impact of two different types of training, both of which used physical balance beam systems [5]. The study compared participants’ pre- to post-test performance on questions about torque and center of gravity following a one-on-one training that involved balancing a meter stick with and without one or two weights either on a fulcrum (a traditional hands-on training) or directly on the participant’s extended fingers (an embodied training). Participants with training showed significant improvement on overall accuracy from pre- to post-test compared with participants in a control group. For the overall accuracies and accuracies on the set of more challenging test questions, the embodied

group outperformed the hands-on group, but these differences in performance were not statistically significant, so that no conclusive group differences were observed.

We have revised our methods to address several weaknesses. The data indicate that our participant pool complicates the analysis but also offers an opportunity to consider how different trainings impact two groups of our participants differently. In this paper, we examine gains in accuracy for participants with lower and higher reasoning abilities who assume one of three roles typically played by a student participating in a “hands-on,” active engagement activity exploring concepts of torque and center of gravity.

II. BACKGROUND

Introductory physics courses are taught for physics and engineering majors, students in the health sciences, or non-STEM students. By design, the level of math applied to similar topics varies greatly across the different courses. Moore and Rubbo showed that STEM and non-STEM students differ significantly in their scientific reasoning ability [6], and not simply in their math skills. Students who enter a physics, astronomy, or general science course with poor scientific reasoning ability are likely to exhibit lower normalized gains on concepts tests than students with high scientific reasoning ability, particularly in areas where the content is categorized as theoretical (e.g. forces or energy). In addition, pedagogies that work for students of high scientific reasoning ability may not work well for students with poor scientific reasoning ability [6].

Lawson’s Classroom Test of Scientific Reasoning (LCTSR) quantifies and categorizes different aspects of student reasoning. The 2000 version of the LCTSR [2] includes 12 scenarios with associated pairs of multiple

choice questions that typically ask for a prediction and then reasoning to support the prediction. Moore and Rubbo found that non-STEM students averaged 54% compared to 75% for STEM students, with the poorest performance (25%) on questions involving proportional reasoning [6]. This result is relevant for our center of gravity training because participants use proportional reasoning to make predictions when they apply torque in the training and solve test questions. While ref. [6] focuses on students in conceptual physics and astronomy courses, algebra-based introductory physics students display a very wide background of mathematical and scientific reasoning abilities. It is possible that our training may impact or benefit students with different abilities in different ways.

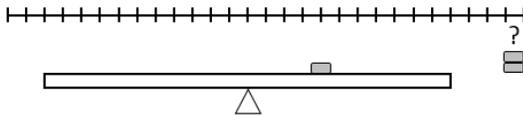
A study by Kwon *et al.* [3] compared the impact on accuracy in solving proportional reasoning problems for teenage male students who used either physical manipulatives or verbal tutoring as part of four 30-minute individual training sessions, one of which used a balance beam task. Students using the physical manipulatives made predictions, received tutoring, and tested their predictions. Students in the verbal group followed the same procedures, but were not allowed to manipulate the materials to test their predicted answers, and “[o]nly pencil and paper and verbal instructions were used” in the tutoring [3]. Participants in the manipulative group significantly outperformed participants in the verbal group on questions requiring proportional reasoning [3].

III. CENTER OF GRAVITY STUDY

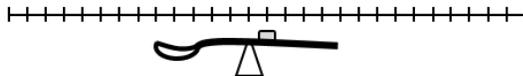
A. Methods

We modified our previous center of gravity study [5] for

(a) Where should a set of blocks that is twice the mass and weight of the first block be placed so that the beam remains balanced?



(b) The disk has the same weight and mass as the spoon shown below. Where is the center of gravity of the spoon?



(c) The rectangle is one-half the mass and weight of the circle. Where does the system of the rectangle and circle balance?

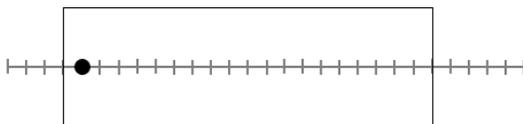


FIG 1. Sample pre- and post-test questions from (a) Category 3, (b) Category 4, and (c) Category 5.

the current study to make the training protocols, and pre- and post-tests more concise, and added an observation group. A test of 36 questions was created. Four sets (Categories 1-4) of eight questions probe participants’ ability (i) to distinguish the center of gravity of an object from the center of gravity of a system, (ii) to apply the principle that an object behaves as if all of its weight is concentrated at its center of gravity, (iii) to apply the concept of torque to simple balanced systems, and (iv) to apply torque and center of gravity to balanced systems. Category 4 questions draw from the more difficult questions from the original study [5]. A fifth set of four questions, Category 5, requires participants to apply torque and center of gravity to systems of a 2-dimensional object and a point object. Samples of questions in Categories 3, 4 and 5 are shown in Fig. 1. Question orders on the pre- and post-tests were different.

Data were collected at two sites: on the campus of the University of Chicago where participants were primarily non-STEM undergraduate students, and at a downtown Chicago location where participants were from the general population. The university group ($N = 52$; 26 male) had an average age of 20.4 yrs, and a median age of 20 yrs. The non-university participants ($N = 59$; 29 male) had average and median ages of 32.3 yrs and 29 yrs, respectively, quite different from the university participants.

Participants in the study were randomly assigned to either the control group or one of three training groups. The hands-on and embodied groups received training similar to that described briefly in Sec. I and in ref. [5]. The observation group was guided through a similar

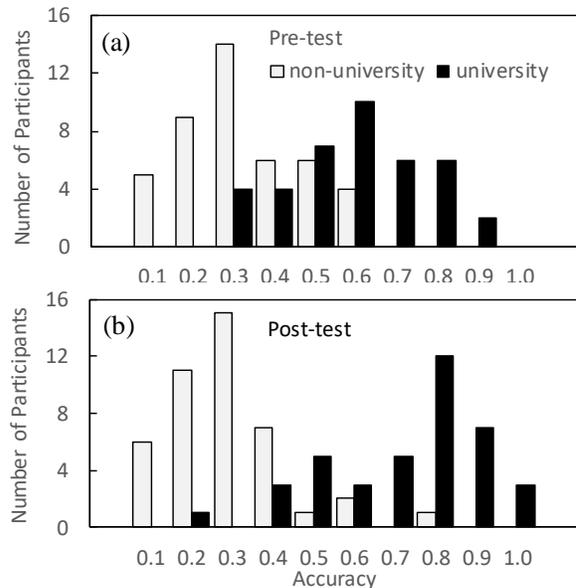


FIG 2. Histograms of accuracies on (a) the pre-test, and (b) the post-test for participants with training at the university (black bars) and the non-university (light grey bars) sites.

training that involved predictions and experimental tests of predictions, but the observer watched as the experimenter handled the meter stick and disks in the same way as a participant from the embodied group. As in the original study, participants in the control group worked on a word search problem between taking the pre- and post-tests.

B. Impact of trainings on different learning populations

Figure 2 shows the distributions in pre-test (upper chart) and post-test (lower chart) accuracies for participants with any of the three types of training from the university and non-university sites. The university participant group shows noticeably higher accuracy at pre-test than the non-university participant group, with the latter group’s distribution changing only marginally from pre- to post-test. Training benefits the university group, which shows a clear shift toward higher accuracy at post-test and no evidence of regression to the mean. The bimodal distribution is suggestive of participants with lower and higher scientific reasoning abilities, similar to the non-STEM and STEM students in ref. [6].

We define two participant pools based on overall pre-test accuracies rather than by site location, where the low accuracy pre-test group have scores below 0.45 ($n = 52$) and the high accuracy pre-test group have scores of 0.45 or greater ($n = 59$). Motivated by the proportional reasoning component of the Lawson test [2], and the study by Kwon *et al.* [3], we first consider how the two population groups perform on our Category 3 questions from pre- to post-test. Both tests included the same eight questions, allowing us to calculate a normalized gain, g , for each participant for these static equilibrium/torque questions that rely on the use of proportional reasoning. Table I describes the method for

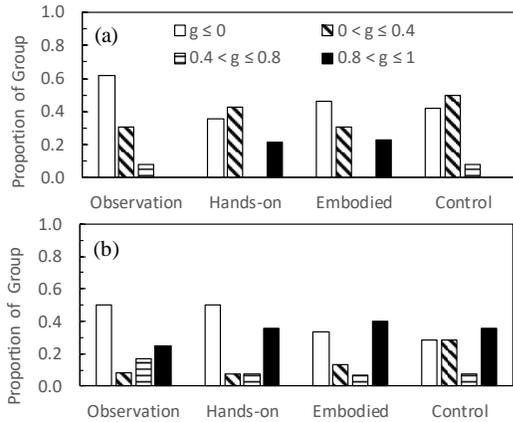


Fig. 3: Proportions of participants with overall pre-test accuracies (a) < 0.45 and (b) > 0.45 in 4 ranges of normalized gains, g , by training group for Category 3 questions.

TABLE I. Formulae for calculating normalized gains for each participant. Categories 3 and 4 questions have $\max = 8$, while Category 5 questions have $\max = 4$.

Case	Gain equation or value
$\text{post} \geq \text{pre}; \text{pre} \neq \max$	$(\text{post} - \text{pre})/(\max - \text{pre})$
$\text{post} = \text{pre} = \max$	data noted but excluded

calculating g for different cases of pre- and post-test scores.

Figure 3 shows normalized gain data for participants in the low-accuracy pre-test population (upper panel) and the high-accuracy population (lower panel). The data are separated according to training group, with 4 possible ranges of normalized gain, as denoted in the figure, for each training group. The lowest range includes negative and zero gains, while the highest range includes a subset of participants who achieved near maximum, along with absolute maximum gain. For Category 3 questions, 2 participants had maximum accuracy at pre- and post-test and are excluded from the data presented. Limited statistics do not allow for meaningful quantitative comparisons, yet are suggestive of the following: For the low-accuracy pre-test population, only participants with hands-on or embodied training achieve maximum or near-maximum gains, while the observation group performs most similar to the control group. For the high-accuracy pre-test population, all groups including the control group, perform similarly, with approximately 1/3 of each group achieving maximum or near-maximum gain. Hands-on and embodied trainings appear to improve the proportional reasoning skills, at least in the context of torque, for the low-accuracy pre-test population, whereas the high-accuracy pre-test population improves simply via the re-testing process.

Category 4 questions require participants to apply

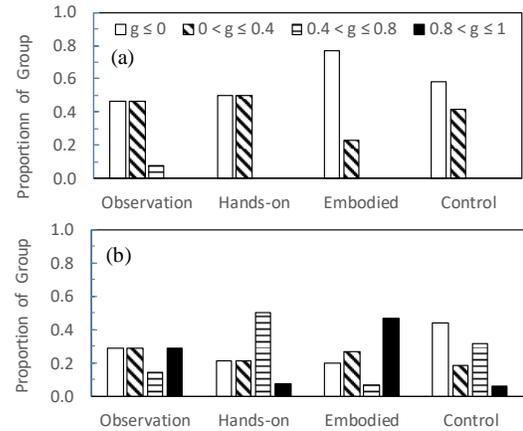


Fig. 4: Proportions of participants with overall pre-test accuracies (a) < 0.45 and (b) > 0.45 in 4 ranges of normalized gains, g , by training group for Category 4 questions.

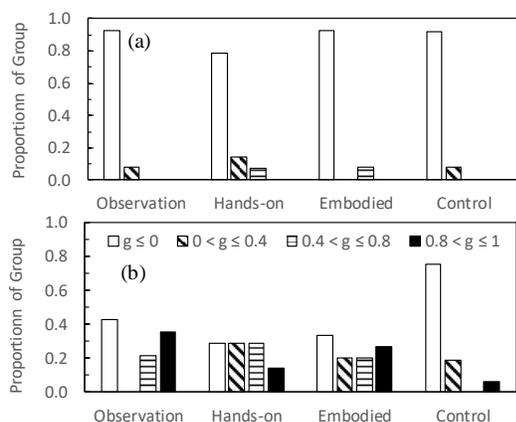


Fig. 5: Proportions of participants with overall pre-test accuracies (a) < 0.45 and (b) > 0.45 in 4 ranges of normalized gains, g , by training group on Category 5 questions.

concepts of torque and center of gravity, thus adding an additional layer of complexity. Data for the four training groups are shown in Fig. 4. Normalized gains are poor for nearly all participants with low overall pre-test accuracies, and gains observed for the torque-only questions do not appear to carry over when problems involve center of gravity in addition to torque. In contrast, participants with high pre-test accuracies show relatively higher gains, with ~ 30 -45% of participants in the embodied and observation trainings achieving maximum or near maximum gain.

Participants apply similar reasoning in the context of 2D shapes for Category 5 questions. See Fig. 5. Data are excluded for 3 participants who had maximum accuracy at pre- and post-test on these questions. Approximately 80% of the participants for every training in the low pre-test accuracy group show no gain from pre- to post-test. For the high pre-test accuracy participants, each training group shows a range of gains, with the observation and embodied groups having the greatest proportion of participants with maximum or near-maximum gain, and all trainings having pronounced, positive impact compared with the control.

IV. IMPLICATIONS AND LIMITATIONS

Data collected for this study are of limited sample size, and while we offer quantitative information, we do not report a statistical analysis. Our training was designed to

compare the performance of participants who received three different types of training that each mimic a role students might play in a “hands-on” classroom activity addressing torque and center of gravity. Participants either balanced a meter stick on a fulcrum or on their hands, or observed an experimenter balance the meter stick on her hands. Each training included making and testing predictions. In a related study [3], students who used physical manipulatives in a balance beam task had significantly higher accuracies on tests of proportional reasoning than students who received verbal tutoring. Similarly for our study, participants with low overall accuracies at pre-test who received hands-on or embodied training showed higher gains on torque only (i.e. proportional reasoning) questions than those in the observation or control groups. Participants who had high overall accuracies at pre-test showed similar gains for all trainings and for the control group, indicating improvement occurs upon repeated questioning. The low-accuracy participants showed low gains for all training conditions when asked to combine proportional reasoning (torque) with the concept of center of gravity. Interestingly, for these more difficult torque + center of gravity questions, along with the 2D transfer questions, the high-accuracy participants with observation and embodied training show higher gains than the hands-on training. We note that our observation training is designed to be identical to the embodied training but with the lack of direct feeling of the balanced system. This is quite different from the verbal instruction group in ref. [3], where pencil and paper took the place of testing with the physical manipulative.

Perhaps more striking is how the low pre-test accuracy and high pre-test accuracy groups respond differently to the different trainings, though the limited size of the data set makes it impossible to state quantitative results. Collecting Lawson (LCTSR) data for our algebra-based introductory physics students would be useful to help us to develop adequate scaffolding on proportional reasoning within the context of torque so that students can successfully apply torque as a tool for exploring center of gravity.

V. ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation EHR Core Research grant DUE-1348614.

[1] K. Pine *et al*, *Devel. Psych.* **40**, 1059 (2004).
 [2] A.E. Lawson *et al.*, *J. Res. Sci. Teach.* **37**, 81 (2000).
 [3] Y.-J. Kwon *et al.*, *J. Res. Sci. Teach* **37**, 1171 (2000).
 [4] P. Marshall *et al.*, in *Proceedings of the 4th International Conference on Tangible, Embedded and Embodied Interaction*, p. 153 (2010).

[5] E. Agra *et al.*, in *2016 PERC Proceedings*, p. 24 (2016).
 [6] J.C. Moore and L.J. Rubbo, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010106 (2012).