

Transferring Transformations: Learning Gains, Student Attitudes, and the Impacts of Multiple Instructors in Large Lecture Courses.

Steven J. Pollock

Department of Physics, University of Colorado, Boulder, CO 80309-0390

Abstract. We have implemented several research-based transformations in our introductory calculus-based physics course at CU Boulder. These include Peer Instruction with student response system in lecture, Tutorials with trained undergraduate learning assistants in recitations, and personalized computer assignments. In an effort to distinguish the effects of instructor, TA preparation, and particular research-based activities, we present extensive new measurements from six courses representing a spectrum of reforms. This study includes data from Physics I with and without Tutorials, and Physics II with Tutorials. We present multiple quantitative and qualitative measures of success, including validated pre/post content- and attitude-surveys and common exam questions. We investigate the hand-off of reforms between faculty implementing different suites of activities, and begin to assess elements and requirements for success with these transformations. We present evidence that *combining* research-based interactive engagement methods in lecture, Tutorials, and homework plays a significant positive role in conceptual and attitudinal development.

Keywords: replication, course assessment

PACS: 01.40.-d, 01.40.Di, 01.40.Gm

INTRODUCTION

When examining the implementation of research-based course transformations, there are at least two central elements to consider. The transformation must be *successful*, based on measurable outcomes consistent with local course goals, and institutional or broader norms. They should also be *replicable and sustainable*: transformation should be reproducible between and within institutions and faculty, without placing undue demands on resources. At our institution (CU Boulder) we have begun a process of transforming our introductory calculus-based sequence, adding Tutorials[1], Peer Instruction with electronic personal response systems[2], and computer homework systems. The goals of these classes include the development of conceptual learning, and constructive beliefs about the nature of learning physics.

In previous work[3], we investigated essential features of these multiple reforms and the dynamics between them in a pair of Physics I courses taught by the author. However, courses are almost completely independent from semester to semester; local culture neither demands nor encourages departmental consensus beyond a rough (traditional) "coverage" of

standard content. Here, we present new data to begin the process of addressing some of the central questions described above: What does it mean to replicate a transformation? In what ways are the transformations successful? What elements of transformation are necessary, which are sufficient, to achieve our goals? As a start, we take an approach of extensive data collection across time, using multiple measures, with over 2000 students in 6 courses (with 6 instructors, some team teaching) over two years.

This work contains two parts: first, we document an explicit attempt at replication of a Physics II course, allowing a number of minor changes, but with an effort made to constrain the changes to maintain the spirit and essential activities of the transformed curriculum. We demonstrate here that we achieve a high level of replication. Next, we document the evolution of Physics I when subsequent instructors did *not* pay specific attention to replication, following their instincts with less direct input from research-based reforms. We demonstrate the successive deterioration of student performance in these Physics I courses. This will lead to future work attempting to understand the role of tools and culture at the level of students and faculty to help explain these results[4].

REPLICATING TRANSFORMATIONS

In Fall '04, the author taught Phys II (445 students in two lectures) using Tutorials[1,5]. Lectures used Concept Tests[2] with peer discussion, and connected explicitly with Tutorials. Tutorial homework was assigned weekly, as well as traditional[6] end-of-chapter problems using a computer system [7]. Exams were 75% multiple choice, 25% long-answer questions based on the Tutorial Instructor's guide[1]. Many students had taken Phys I with Tutorials. Grad TAs, teamed with undergrad Learning Assistants (LAs) [5], attended weekly prep sessions following the University of Washington model.

In Sp '05, the Phys II course was handed off to a different instructor. Most of these students came from a Phys I course which had not used Tutorials. They used a new textbook[8]. Lectures used Concept Tests and peer discussion, with many of the same questions from the previous term. Exams were strictly multiple choice, with a heavy emphasis on conceptual understanding. The weekly Tutorial preparatory meetings were structured the same as the previous semester, although grad student observers from our PER group noted that there was a slightly larger emphasis on content, rather than pedagogical training.

To compare the courses, we collected four types of data. A pre/post conceptual exam (BEMA[9]) was administered in the first and last weeks of the term. A pre/post research-based survey on student attitudes and beliefs about learning (CLASS[10]) was administered online. Each exam had several questions from the previous term, for direct comparison on a broader spectrum of topics than the BEMA covers. Finally, informal surveys were given online to assess students' opinions about the value of course elements.

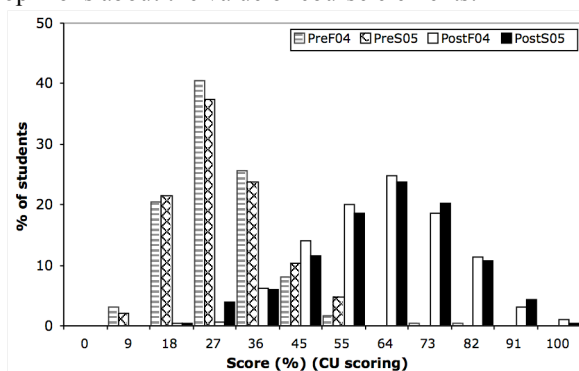


FIGURE 1. Pre and post BEMA results for two courses. Fa04 pre average is $26 \pm 1\%$, ($N=319$ matched), st. dev= 10% . Sp05 pre average is $27 \pm 1\%$ ($N=232$ matched). Both post averages are $59 \pm 1\%$. Both courses match each other closely.

Fig 1 shows a comparison of BEMA results for the two courses. Despite having somewhat different demographics, both courses had statistically identical

pre and post scores. This normalized learning gain (44%) is not as high as we see in mechanics, but is quite high compared to traditional instruction[11]. This exam is difficult; grad TA's started the term with an average score of 82%. Volunteers just completing our upper division E&M sequence averaged 52%, *below* our average freshman posttest score.

Over four exams, we asked 31 common multiple-choice questions, 13 of which were based explicitly on Tutorials. The difference in average scores for all questions, as well as for the subset of 13, was $<2\%$ (not statistically significant). On the CLASS, both classes saw very small (although statistically significant) declines in "expert-like" beliefs of $\sim 2-3\%$ (from 66-68% to 64-65%). With traditional instruction, larger regression ($\sim 10\%$) is more typical[10,12]. On informal surveys, in which students were asked Likert-scale questions like "The Tutorial helps me learn physics" and "I enjoyed the Tutorial", the students were fairly neutral. There was a very small trend towards more negative response in the second implementation. (For the "enjoyed Tutorials" question, on a scale of 1-5 with 5="strongly agree", the Fa04 average was 3.3 ± 1 , compared to 3.0 ± 1 in Sp05.)

By all these measures, the course outcomes were virtually indistinguishable. The effects of different semester, instructor, textbook, and exam style were not large enough to cause measurable shifts in pre/post and exam performance. The choice of nearly identical content coverage, common use of peer instruction and specific concept questions in lecture, Tutorials in recitations, and the general emphasis on conceptual understanding apparently dominated.

EVOLUTION OF TRANSFORMATIONS

In Fall '03 and Sp '04, the author taught Phys I, using Tutorials. The implementations of these two courses were nearly identical to each other, and to the Fa '04 Phys II described in the previous section. The FCI [13] was used in Fa '03, the FMCE[14] in Sp '04. Results from these two Phys I courses matched closely on exams, pre/post CLASS, and affect surveys, with extremely high learning gains on conceptual surveys. [3,15] For comparison to the courses which followed, we combine exam results from these two terms, and use the Sp '04 FMCE scores as a basis of comparison. We refer to this as implementation "#1" below.

In Fa '04, Phys I was taught by two instructors, both of whom are popular and well-rated teachers. The lead instructor was an early adopter of PER-based reforms, and regularly spoke with PER-research faculty. They used a nominally research-based text[16], using Concept Tests in lecture, supplied with

that text. Class observers noted that peer discussion was more muted in this course than in the previous semesters. The instructors also switched to a research-based homework system[8]. The secondary instructor opted not to use Tutorials in recitations. He did encourage TA's to put students into small groups, using the text's Student Workbook[16]. This "workbook" format was followed for almost half the semester, when recitations reverted to more traditional homework review sessions, still with group work, due to a few vocal complaints from students. There was little support or training of the TA's after the first several weeks of the term, and no undergrad LA's [4,5] were used. We call this implementation "#2" below.

In Sp '05, the secondary instructor from Fa '04 led the course, with a new instructor assisting on background tasks. Concept Tests were used in lecture, but the new instructor made up his own questions, with little overlap from previous terms. Observers and the instructor noted that peer discussion was more muted than the previous term, although still present. Recitations were traditional, with TA's answering questions about homework. Several returning TA's encouraged occasional group work, centered on homeworks. This approach was "implementation #3".

To compare these courses, we collected three types of data: the FMCE (pre/post in recitation), the CLASS (pre/post online), and common exam questions. The instructors for implementations #2 and #3 freely chose exam questions from implementation #1. About 30 questions each term allowed direct comparisons between the implementations. The instructors for #2 and #3 chose not to give survey questions on student affect.

To summarize the variation of implementation of Physics I: all courses used Concept Tests, but with decreasing emphasis and support of student discussion in lectures. Courses #2 and #3 both used the same text [15], and the same homework system [8], each of which was research-based (unlike those used in #1). The most significant structural difference was the use of Tutorials in #1, with trained TAs and LAs. Small group activities[16] were used in #2 with untrained TA's, and traditional recitations were used in #3. All 3 courses should be characterized as "transformed", but with a systematic decrease in the amount and level of research-based activities in lecture and recitation.

The pre-test FMCE score in course #2 was 5% higher than in the others, a statistically significantly higher starting group. (This was the only Fall course, with different demographics, including 30 physics majors compared to <10 in either Spring course.) Fig. 2 shows the pre/post FMCE results, binning by normalized learning gain. Ave learning gains successively declined each semester, from 66% to 59% to 45% (all $\pm 2\%$).

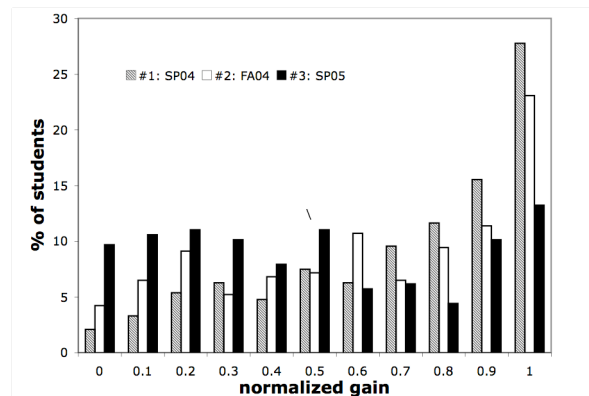


FIGURE 2. Histogram of normalized learning gain on FMCE for course #1 (grey), 2 (hollow), and #3 (solid). Pretest scores were, in order, $28 \pm 1\%$ ($N=335$, matched, st. dev. = 20%), $33 \pm 1\%$ ($N=302$) and $28 \pm 1\%$ ($N=213$). Posttest scores were $74 \pm 1\%$, $69 \pm 2\%$, and $59 \pm 2\%$.

It is interesting to ask what the impact of the different implementations was on students with different starting scores on the FMCE. Fig. 3 shows high learning gains for students with high incoming pretests (>50%) for implementations #1 and #2 (with and without Tutorials), but significantly lower gains for this population in the class with traditional recitations. (We have found some correlation between starting score and learning gain [3] in the past, so one might expect slightly *higher* learning gains for course #2.) For students with low incoming scores (<15%), implementation #1 shows the highest gains, with #3's gain the lowest, although still comparable to (good) traditional instruction[15].

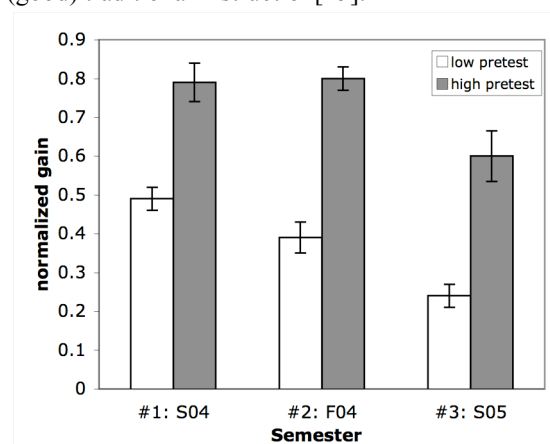


FIGURE 3. Impact on different student populations. White shows ave normalized gain on FMCE for students *entering* with very low (<15%) pretest scores (These represent $\sim 20\%$ of the students in each class). The shaded "high pretest" population accounts for <15% of students in courses #1 and #3, 22% of course #2.

Over four exams, implementation #2 asked 34 multiple-choice exam questions in common with implementation #1. The decline in average scores for these questions from implementation #1 to #2 was ~6 points (out of 100), a statistically significant difference ($p < .01$, two-tailed paired t-test). On "Tutorial-based" questions (half these common questions), the drop was ~7 points. On more traditional (quantitative) problems (one quarter of the questions), the difference was ~5 points. Comparing course #1 to #3 we find even larger differences. Of 30 common exam questions, the average score in course #3 was ~9 points lower. On "Tutorial-based" questions the decline was ~13 points. On computational (traditional) common questions, the drop was ~8 points. In these classes, 12 points is generally a full letter grade difference. Thus, performance on a broad spectrum of exam questions consistently indicate a steady decline in student learning across this sequence of courses.

On the CLASS, implementation #1 saw very small increases in "expert-like" beliefs over the course of the term (shift of $+1.5 \pm 1\%$) However, in courses #2 and #3, there were *declines* over the course of the semester more in keeping with traditional instruction[10,12], with shifts in "expert-like" opinions of $-6.5 \pm 1\%$.

CONCLUSIONS

We would like to better understand the necessary and sufficient conditions for transferring, replicating, and sustaining successful learning environments. It is common to assess transformed curricula by adding an experimental method to a control group, and then measuring the impact. In our case, we take a slightly different approach - implementing a large number of research-based innovations[3] and observing the impact as faculty change or remove one or more elements in the mix. Although not a controlled experiment, one can study simultaneously the effectiveness of the reforms, and address the replicability and sustainability in a variety of realistic circumstances. We see that some instructors can successfully appropriate multiple elements in a very large class situation with modest resources [3-5].

We consistently find that no single reform element appears to uniquely account for the measured successes[3], but the combination of course elements has a large impact. In one case, changing instructor, along with text and exam structure, did not lead to significant changes in learning gains, exam results, or student attitudes. However, when left to naturally evolve with no oversight, we observed a trend towards more traditional implementations and declining measures of student learning[17]. The downward shift on all measures from implementation #2 to #3, where

the most significant change in course structure was in the recitations, indicates the importance of interactive engagement in these environments. These data provide a quantitative groundwork with which to contribute to deeper questions raised by this study - why do we see worse student performance in some environments, why do professors choose which reforms to implement and how do they go about doing so, and what are the connections between class norms and culture, and the impact on students?

ACKNOWLEDGMENTS

Thanks to APS Colorado PhysTEC and NSF CCLI (DUE #0410744) for grants which support our transformed classes, and the NSF STEM-teacher preparation grant which supports learning assistants. Enormous thanks also to Noah Finkelstein, the University of Washington's Physics Education Group, CU Physics, and the whole PER at Colorado group.

REFERENCES

1. McDermott, L., Shaffer, P., and the PEG, *"Tutorials in Introductory Physics"*, Prentice Hall 2002.
2. Mazur, E., *Peer Instruction*, Prentice Hall 1997
3. Pollock, S., *"No single cause..."*, PERC Proceedings 2004, Sacramento, CA.
4. Finkelstein, N. Pollock, S., *"Replicating&Understanding Successful Innovations"*, Phys. Rev. ST Phys. Educ. Res. 1, 010101 ('05).
5. Finkelstein, N. et al., PhysTEC 2005 Annual Summary Report, see phystec.colorado.edu
6. Halliday, Resnick, Walker, *Fundamentals of Physics* Wiley 2004.
7. CAPA homework system: see www.lon-capa.org
8. Mastering Physics: see www.masteringphysics.com
9. Brief Electricity and Magnetism Assessment, see <http://www.ncsu.edu/per/TestInfo.html>
10. <http://cosmos.colorado.edu/phet/survey/fall2003.html>. W. Adams et al., *"Design and Validation of the Colorado Learning About Science Survey"*, and K. Perkins et al., *"Correlating student attitudes with student learning using the CLASS"*, both in PERC Proceedings 2004, Sacramento, CA.
11. Chabay, R. and Sherwood, B., private communication.
12. Redish, Saul, Steinberg *Am. J. Phys.* **66** 212-224, (1998)
13. Hestenes, D. Wells, M., and Swackhamer, G., Force Concept Inventory. *Physics Teacher* **30** (1992), 141-158
14. Thornton, R., Sokoloff, D., *Am. J. Phys* **B66**(4), 338-352, 1998, also Thornton, R., Kuhl, D., Cummings, K., Marx J., *"Comparing the FMCE and the FCT"*, unpublished
15. Hake, R., *American Journal of Physics* **66**, 64-74, (1998) Hake, R., *Conservation Ecology* **5**(2):28, (2002)
16. Knight, R., "Physics for Scientists and Engineers", with accompanying "Student Workbook". Pearson 2004.
17. Henderson, C., and Dancy, M, PERC Proceedings 2004.