

Compressed-format compared to regular-format in a first-year university physics course

Jason J.B. Harlow, David M. Harrison,^a and Eli Honig

Department of Physics, University of Toronto, Toronto, ON M5S 1A7

We have compared student performance in two sessions of a large first-year university physics course, one with a normal 12-week term and the other with a compressed 6-week term. Student performance was measured by the normalized gain on the Force Concept Inventory. The gains for the regular-format course were better than the gains for the compressed-format course, and while the differences in gains are small they are statistically significant. Not accounted for are the differences in effectiveness of the different instructors in the two versions of the course.

I. INTRODUCTION

At the University of Toronto our first year Physics course intended primarily for students in the life sciences is PHY131. We have compared the performance of PHY131 students in the regular fall 12-week term to a compressed 6-week format in the summer term: the fall version was given in 2012 and the summer version in 2013. There is also a 12-week version of the course given in the winter term, which is not part of this study. In addition, there is a separate first-year course for physics majors and specialists, and another for engineering science students: neither of these courses are part of this study.

We measured student performance using the Force Concept Inventory (FCI). The FCI has become a common tool for assessing students' conceptual understanding of mechanics, and for assessing the effectiveness of instruction. It was introduced by Hestenes, Wells, and Swackhamer in 1992,¹ and was updated in 1995.² The FCI has now been given to many thousands of students at a number of institutions worldwide. A common methodology is to administer the instrument at the beginning of a course, the "Pre-Course," and again at the end, the "Post-Course," and look at the gain in performance. Our students were given one-half a point, 0.5%, towards their final grade in the course for answering all questions on the Pre-Course FCI, regardless of what they answered, and given another half point for answering all questions on the Post-Course FCI, also regardless of what they answered. Below all FCI scores are given as a percent.

a. Electronic mail: david.harrison@utoronto.ca

PHY131 is the first of a two-semester calculus-based sequence, and the textbook is by Knight.³ A senior-level high-school physics course (“Grade 12 Physics”) is recommended as a pre-requisite but is not required to take this course. One of us (JJBH) was one of the two lecturers in the fall session, and DMH was the sole lecturer in the summer session. In the fall session the two lecturers alternate so all students are given instruction by the same lecturers. Research-based instruction is used throughout the course; clickers, Peer Instruction,⁴ and Interactive Lecture Demonstrations⁵ are used extensively in the classes. In the fall term there are two one-hour classes every week, while the summer course has two 2-hour classes per week in a single session. In addition, due to logistic constraints the summer course has a total of 22 hours of classes, while the fall version has 24 hours of classes.

Traditional tutorials and laboratories have been combined into a single active-learning environment, which we call *Practicals*;⁶ these are inspired by Physics Education Research tools such as McDermott’s *Tutorials in Introductory Physics*⁷ and Laws’ *Workshop Physics*.⁸ These Practicals are similar in many ways to the *Labatorials* at the University of Calgary.⁹ In the Practicals students work in teams of four on conceptually-based activities using a guided-discovery model of instruction. Whenever possible the activities use a physical apparatus or a simulation. Some of the materials are based on activities from McDermott and from Laws. In the fall term there is one 2-hour Practical every week, and the summer course has two 2-hour Practicals per week.

II. METHODS

The FCI was given during the Practicals, the Pre-Course one during the first week of classes and the Post-Course one during the last week of classes. For both courses, over 95% of the students who were currently enrolled took the FCI. For the Fall 2012 session, 868 students took the Pre-Course and 663 students took the Post-Course; for the Summer 2013 session 239 students took the Pre-Course and 193 students took the Post-Course. The students who took the Pre-Course but not the Post-Course FCI were almost all students who had dropped the course. These dropout rates, 24% for the fall session and 19% for the summer one, are typical for this course.

There is a small issue involving the values to be used in analyzing both the Pre-Course and the Post-Course FCI numbers. Below we will only analyze the data for students who took both the Pre-Course and Post-Course FCI—the “matched” values—which consisted of 641 students for the fall term and 190 for the summer term. The students who took the Post-Course FCI but not the Pre-Course FCI were probably either late enrollees or they missed the first Practical for some other reason. In all cases, the difference in median FCI scores between using raw data or matched data is only a few percent. These small differences between matched and unmatched data are consistent with a speculation by Hake for courses with enrollments of greater than 50 students.¹⁰

Figure 1 shows the FCI scores for the fall term and Fig. 2 shows the scores for the summer term. None of the histograms conform to a Gaussian distribution, especially for

the Post-Course results. Therefore, characterizing the results by the mean value is not appropriate, and the median value will provide a better measure of central tendency and is what we will use below, although for our data the differences in FCI scores between the means and medians is fairly small, between 0.8% and 2.6%. The uncertainty in the median is taken to be the inter-quartile range divided by \sqrt{N} .¹¹

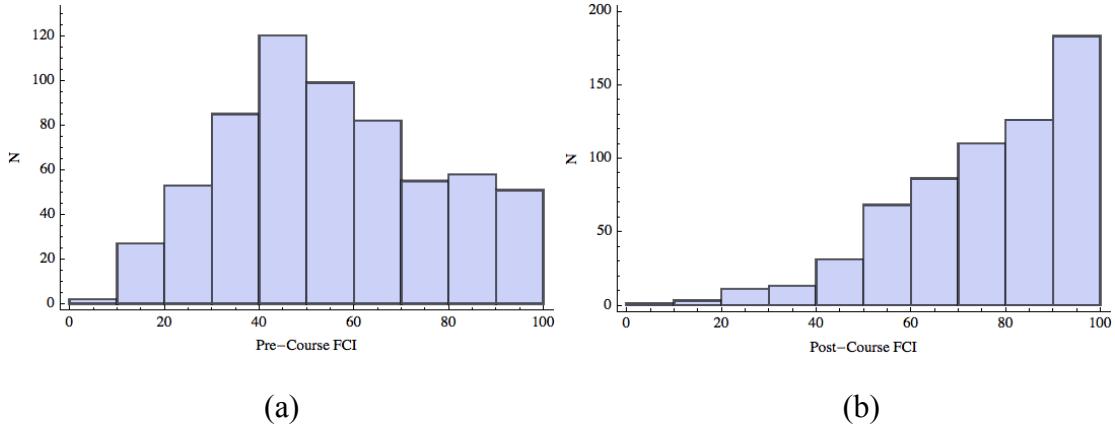


Figure 1. Histograms of the Pre-Course and Post-Course FCI Scores for the fall session

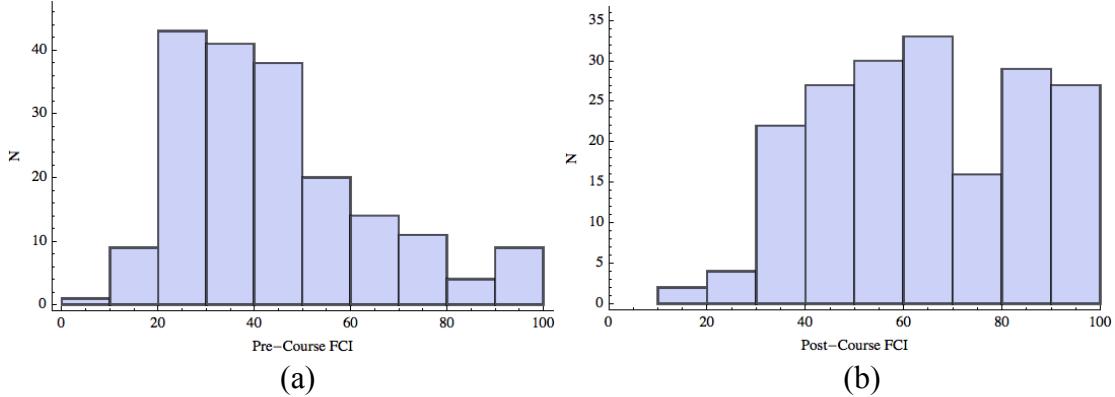


Figure 2. Histograms of the Pre-Course and Post-Course FCI Scores for the summer session

We also calculated gains on the FCI. The standard way of measuring student gains is from a seminal paper by Hake.¹² It is defined as the gain normalized by the maximum possible gain:

$$G = \frac{\text{Post \%} - \text{Pre \%}}{100 - \text{Pre \%}} \quad (1)$$

Clearly, G cannot be calculated for Pre-Course scores of 100, which was the case for nine students in the fall term and none in the summer.

One hopes that the students' performance on the FCI is higher at the end of a course than at the beginning. The standard way of measuring the gain in FCI scores for a class is called the average normalized gain, to which we will give the symbol $\langle g \rangle_{\text{mean}}$, and was also defined by Hake in Ref. 12:

$$\langle g \rangle_{\text{mean}} = \frac{\langle \text{Post\%} \rangle - \langle \text{Pre\%} \rangle}{100 - \langle \text{Pre\%} \rangle}, \quad (2)$$

where the angle brackets indicate means. However, as discussed, since the distribution of FCI scores is not Gaussian, the mean is not the most appropriate way of characterizing FCI results. We will instead report $\langle g \rangle_{\text{median}}$, which is also defined by Eq. (2) except that the angle brackets on the right hand side indicate medians instead of means. The normalized gains are often taken to be a measure of the effectiveness of instruction.

The uncertainties in the average and median normalized gains reported below are the propagated errors in the Pre-Course and Post-Course FCI scores. Since both of these are errors of precision, they should be combined in quadrature, i.e. the square root of the sum of the squares of the uncertainties in the Pre-Course and Post-Course scores. Therefore,

$$\begin{aligned} \Delta(\langle g \rangle) &= \sqrt{\left(\frac{\partial \langle g \rangle}{\partial \langle \text{Pre\%} \rangle} \Delta \langle \text{Pre\%} \rangle \right)^2 + \left(\frac{\partial \langle g \rangle}{\partial \langle \text{Post\%} \rangle} \Delta \langle \text{Post\%} \rangle \right)^2} \\ &= \sqrt{\left[\frac{\langle \text{Post\%} \rangle - 100}{(\langle \text{Pre\%} \rangle - 100)^2} \Delta \langle \text{Pre\%} \rangle \right]^2 + \left[\frac{\Delta \langle \text{Post\%} \rangle}{100 - \langle \text{Pre\%} \rangle} \right]^2}, \end{aligned} \quad (3)$$

where $\Delta \langle \text{Pre\%} \rangle$ and $\Delta \langle \text{Post\%} \rangle$ are the inter-quartile ranges divided by \sqrt{N} . In addition to the FCI, we collected data on student background and why they were taking the course. In the fall session we collected that data using clickers in the second week of classes: only about 500 of the 641 “matched” students answered these questions. In the summer session we avoided this unfortunate loss of 22% of the sample by including the questions as part of the Pre-Course FCI.

III. RESULTS

From Figs. 1 and 2, we can see that the summer students begin and end the course with less conceptual understanding of classical mechanics than their counterparts in the fall session. In addition, the FCI results are also reported in Table I, along with the median normalized gains. The table shows that in terms of $\langle g \rangle_{\text{median}}$, the fall session outperformed the summer session by 0.14 ± 0.06 . The value of 0.14 is roughly double the uncertainty of 0.06, so the value is significant at roughly a 95% confidence interval.¹³ Calculating $\langle g \rangle_{\text{mean}}$ as is standard in the literature gives values of 0.45 ± 0.02 for the

fall session and 0.34 ± 0.03 for the summer session, where the uncertainties are the propagated standard errors of the mean, σ / \sqrt{N} . These values are both consistent with those of first year university courses using reformed research-based pedagogy, as can be seen, for example, in Ref. 12.

Earlier we stated that the difference between using “matched” students who took both the Pre-Course and Post-Course FCI compared to using all the FCI scores was only a few percent. As an example, for the summer 2013 session the median normalized gain using all FCI scores was 0.42 ± 0.04 .

Another way to investigate the difference in gains in the two sessions is to look at the individual normalized gains G . Figure 3 shows histograms of the gains for the fall term (panel a) and the summer term (panel b). The median values for the G values are given in the final row of Table I. The difference in the median values of G is 0.15 ± 0.04 , which is about the same as the differences in $\langle g \rangle_{\text{median}}$.

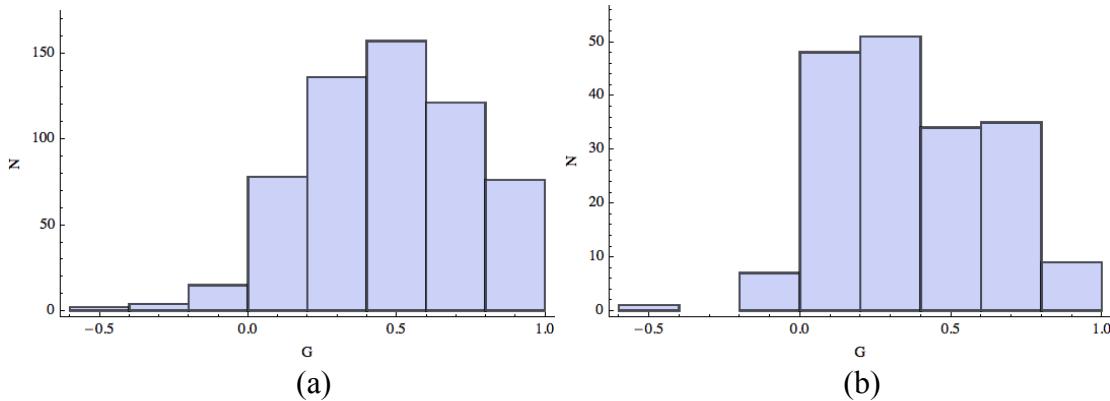


Figure 3. Individual gains G for the fall session (a) and the summer session (b)

An alternative to histograms for visually comparing two distributions is the boxplot. Figure 4 shows the boxplot for the two distributions of G . The “waist” on the box plot is the median, the “shoulder” is the upper quartile, and the “hip” is the lower quartile. The vertical lines extend to the largest/smallest value less/greater than a heuristically defined outlier cutoff.¹⁴ The dots represent data that are considered to be outliers. Also shown in the figure are the statistical uncertainties in the value of the medians.

There were 10 students in the fall session whose Pre-Course score was over 80% and whose G was less than -0.66; one such student had a G of -4 and two had G s of -3. These students were clearly outliers, and probably just “blew off” the Post-Course FCI. There were no such students in the summer session. We have chosen a range for the vertical axis in Fig. 4 so that those 10 students are not shown.

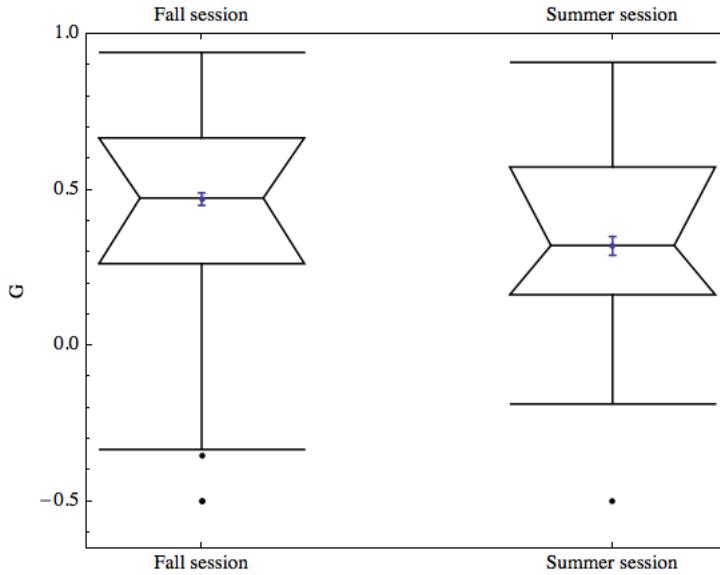


Figure 4. Boxplots for the distributions of G for the two sessions of the course

Table I. Values of FCI scores and gains for the two sessions

	Fall 2012	Summer 2013
Pre-Course FCI Medians	50.0 ± 1.3	40.0 ± 1.9
Post-Course FCI Medians	76.7 ± 1.1	63.3 ± 2.9
$\langle g \rangle_{\text{median}}$	0.53 ± 0.03	0.39 ± 0.05
G Medians	0.47 ± 0.02	0.32 ± 0.03

Although the data show that the overall performance of students in the fall session was better than the performance of the summer students, the question remains as to whether the differences are significant. For data that conforms to a Gaussian distribution, the fairly well known Cohen d effect size is often used to characterize the difference in two distributions.¹⁵ For data like our values for G , which are not normally distributed, Cliff's δ provides a similar measure.¹⁶ The Cliff δ for 2 samples is the probability that a value randomly selected from the first group is greater than a randomly selected value from the second group minus the probability that a randomly selected value from the first group is less than a randomly selected value from the second group. It is calculated as:

$$\delta = \frac{\#(x_1 > x_2) - \#(x_1 < x_2)}{N_1 N_2} \quad (4)$$

where the symbol # indicates counting. The values of δ can range from -1 , when all the values of the first sample are less than the values of the second, to $+1$, where all the values of the first sample are greater than the values of the second. A value of 0 indicates samples whose distributions completely overlap. By convention, $|\delta| \sim 0.2$ indicates a

“small” difference between the two samples, $|\delta| \sim 0.5$ a “medium” difference, and $|\delta| \sim 0.8$ a “large” one.¹⁷ For our G values $\delta = 0.227$ with a 95% confidence interval range¹⁸ of 0.136–0.314. Thus, the difference in G is small, but since the confidence interval range does not include zero it is statistically significant.

For the fall session, the correlation of FCI performance with student interest and background as has been studied.¹⁹ As shown here in the Appendix, except for their program of study (Question 1) the characteristics of the student population in the two sessions of the course are different. Therefore, we were surprised that when each course was examined separately, the normalized gains were the same within the uncertainties for each student characteristic and were also consistent with the overall values of Table I. Table II summarizes the median normalized gains for the two sessions for those characteristics that had a significant number of students in all categories and with different pre-course and post-course FCI scores (some of this data for the fall session also appeared in Ref 19). Also, recall that twice the stated uncertainty corresponds roughly to a 95% confidence interval.

Table II. Normalized Gains for Various Groups

Student Category	Fall Session	Summer Session
	$\langle g \rangle_{\text{median}}$	$\langle g \rangle_{\text{median}}$
Taking the course because it is required	0.47 ± 0.04	0.35 ± 0.05
Taking the course for their own interest	0.56 ± 0.11	0.32 ± 0.15
Taking the course both because it is Required and for their own interest	0.57 ± 0.04	0.41 ± 0.11
Took Grade 12 Physics	0.54 ± 0.03	0.35 ± 0.07
Did not take Grade 12 Physics	0.55 ± 0.04	0.45 ± 0.07
Previously started but dropped the course	0.48 ± 0.20	0.25 ± 0.14
Had not previously started the course	0.59 ± 0.03	0.42 ± 0.05

From Table II, the median Pre-Course FCI score for the fall session is greater than for the summer session by $\Delta = 13.3 \pm 2.3$. As discussed in Ref. 19, fitting G vs the Pre-Course FCI scores for the fall session gave a small but positive slope of $m_{\text{fall}} = 0.00212 \pm 0.00054$. Similar fits for the summer session gave a slope of $m_{\text{summer}} = 0.0018 \pm 0.0021$, which is also positive but is zero within uncertainties. Coletta and Phillips propose that a “hidden variable”—the cognitive level of the students—is responsible for such non-zero slopes.²⁰ Regardless of the cause, such slopes might be expected to give changes in the median normalized G of something on the order of $\Delta \times m_{\text{fall}} = 0.028 \pm 0.009$ or $\Delta \times m_{\text{summer}} = 0.02 \pm 0.03$. The values are less than the uncertainty in the difference in $\langle g \rangle_{\text{median}}$ (0.06) or the uncertainty in the difference in the values of the median of G (0.04) for the two courses. Thus, the effect of the differing Pre-Course FCI scores for the two courses appears to have a negligible impact on the values of the normalized gains.

IV. DISCUSSION

So far as we know, there is only one other study that compares student performance in compressed-format to regular-format in a post-secondary introductory physics course.²¹ This study used the Force and Motion Conceptual Evaluation (FMCE).²² However, in this study the type of pedagogy used in the different sessions of the course was somewhat different, although the instruction was more-or-less based on Physics Education Research in all versions. The results were that the normalized gains on the FMCE were somewhat less for the summer session than for either of the regular sessions that were studied.

There is also a study of summer intensive high-school physics courses for gifted students, which used the FCI and found that the gains were comparable to those of ordinary year-long courses taken by average students.²³ There are studies of post-secondary courses in accounting²⁴, and basic skills in math, English, and reading in U.S. community colleges.^{25,26} The community college studies report that the compressed format is at least as effective as the regular format offerings, although they do not have a tool like the FCI to do a rigorous measurement of the effectiveness of the courses.

Our intuition was that the compressed 6-week format of the summer course does not allow adequate time for students to reflect upon and absorb the sometimes counter-intuitive concepts of classical mechanics; some have compared this version of the course to trying to drink from a fire hose. Despite the fact that the present authors together have nearly 100 years of physics teaching experience, our intuition did not really correspond to the data. Although there are differences in the gains of the students, they are not nearly as dramatic as we had expected. This study, then, is another example of the importance of the methodology of Physics Education Research: teachers' intuition is sometimes wrong but the data do not lie.

The fall session had two lecturers who alternated, and the summer session had a different lecturer. The two courses use almost identical pedagogy both in class and in the Practicals, and the three lecturers communicated regularly and occasionally observed each other's classes. Hoellwarth and Moelter studied instructor effectiveness in a Studio Physics course over 9 quarters with 11 different instructors, and found that there were variations in the values of $\langle g \rangle_{\text{mean}}$ for the same instructor for different quarters that were more significant than the variation between different instructors.²⁷ Thus it is reasonable to assume that the small but statistically significant differences in performance of students in the two sessions studied here are largely due to the different formats of the course. Any differences due to the different effectiveness of the different instructors are perhaps at most second-order effects.

Finally, as stated in Ref. 19, we are continuing to collect this sort of data. We hope that with more students in the sample and better statistics we can revisit the question of normalized gains for different groups and sub-groups of students and the effectiveness of compressed-format compared to regular-format courses. Such longitudinal data will also allow us to get a better idea about the importance of the individual instructor or instructors in influencing student performance.

ACKNOWLEDGEMENTS

We particularly thank our Teaching Assistants, whom we call *Instructors*, whose great work in the Practicals is a vital part of the students' learning. We also thank April Seeley in the Department of Physics for assistance in the collecting of the data of this study. We benefited from discussions about this work with David C. Bailey in the Department of Physics. We also acknowledge two anonymous reviewers for this journal whose comments helped strengthen the paper.

APPENDIX

We asked the students to self-report on the reason they are taking the course and some background information about themselves. Here we summarize that data.

1. "What is your intended or current Program of Study (PoST)?"

Answer	Fall 2012	Summer 2013
Life Sciences	88%	90%
Physical and Mathematical Sciences	9%	8%
Other/Uncertain	4%	2%

2. "What is the main reason you are taking PHY131?"

Answer	Fall 2012	Summer 2013
"Because it is required"	32%	60%
"For my own interest"	16%	12%
"Both because it is required and for my own interest"	52%	28%

3. "When did you graduate from high school?"

Answer	Fall 2012	Summer 2013
2012	78%	41%
2011	9%	20%
2010	5%	21%
2009	3%	11%
Other/NA	4%	7%

4. “Did you take Grade 12 Physics or an equivalent course elsewhere?”

Answer	Fall 2012	Summer 2013
Yes	75%	61%
No	25%	39%

5. “MAT135 or an equivalent calculus course is a co-requisite for PHY131. When did you take the math course?”

Answer	Fall 2012	Summer 2013
“I am taking it now”	81%	5%
“Last year”	10%	53%
“Two or more years ago”	9%	42%

6. “Have your previously started but did not finish PHY131?”

Answer	Fall 2012	Summer 2013
Yes	4%	15%
No	96%	85%

REFERENCES

- ¹ D. Hestenes, M. Wells and G. Swackhammer, "Force Concept Inventory," *The Physics Teacher* **30**(2), 141–158 (1992).
- ² Available from <<http://modeling.asu.edu/R&E/Research.html>>.
- ³ Randall D. Knight, *Physics for Scientists and Engineers: A Strategic Approach*, 3rd ed. (Pearson, Toronto, 2013).
- ⁴ E. Mazur, *Peer Instruction: A User’s Manual* (Addison-Wesley, New York, 1996).
- ⁵ D.R. Sokoloff and R.K. Thornton, “Using Interactive Lecture Demonstrations to Create an Active Learning Environment,” *The Physics Teacher* **35**(6), 340–347 (1997).
- ⁶ The U of T Practicals web site is <<http://www.upscale.utoronto.ca/Practicals/>>.
- ⁷ L.C. McDermott, P.S. Schaffer and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice Hall, New Jersey, 2002).
- ⁸ P.W. Laws, *Workshop Physics Activity Guide* (Wiley, 2004).
- ⁹ D. Ahrensmeir, J.M.K.C. Donev, R.B. Hicks, A.A. Louro, L. Sangalli, R.B. Stafford, and R.I. Thompson, “Labatorials at the University of Calgary: In pursuit of effective small group instruction within large registration physics service courses,” *Physics in Canada* **65**(4), 214–216 (2009).
- ¹⁰ Richard Hake, “Lessons from the Physics Education Reform Effort,” *Ecology and Society* **5**(2), 28 (2002), available online at <<http://www.ecologyandsociety.org/vol5/iss2/art28/>>.

¹¹ This is similar in spirit but simpler than a recommendation in B. Iglewicz, “Robust Scale Estimators and Confidence Intervals for Location,” in D.C. Hoaglin, F. Mosteller, and J.W. Tukey eds., *Understanding Robust and Exploratory Data Analysis* (Wiley-Interscience, Toronto, 1983), p. 424.

¹² R.R. Hake, “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” *Am. J. Phys.* **66**(1), 64–74 (1998).

¹³ If the uncertainties were based on the standard error of the mean, twice the uncertainty’s value would be a 95% confidence interval, but for the median, as discussed in Reference 11, the interpretation is more indirect.

¹⁴ There are various conventions for the cutoff definition. We use 1.5 times the interquartile range extending from the upper and lower quartiles, which was proposed in J.D. Emerson and J. Strenio, “Boxplots and Batch Comparison,” in D.C. Hoaglin, F. Mosteller, and J.W. Tukey eds., Reference 11, p. 58. This cutoff definition is the usual one.

¹⁵ J. Cohen, “A power primer,” *Psychological Bulletin* **112**(1), 155–159 (1992).

¹⁶ N. Cliff, “Dominance statistics: Ordinal analyses to answer ordinal questions,” *Psychological Bulletin* **114**(3), 494–509 (1993).

¹⁷ This way of characterizing the effect size is not without controversy and can be overused. Romano, Kromrey, Coraggio and Showronek go even further and recommend setting thresholds of $|\delta| < 0.147$ “negligible”, $|\delta| < 0.33$ “small”, $|\delta| < 0.474$ “medium”, and “large” otherwise, in “Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and Cohen’s d indices the most appropriate choices?,” presented at the annual meeting of the Southern Association for Institutional Research, October, 14–17, 2006, Arlington, Virginia.

[\(Retrieved September 29, 2014\).](http://www.coedu.usf.edu/main/departments/me/documents/MethodsforEvaluatingGroup.pdf)

¹⁸ The lower value is the 95% CI *infimum* and the upper value is the *supremum*.

¹⁹ J.J.B. Harlow, D.M. Harrison, and A. Meyertholen, “Correlating student interest and high school preparation with learning and performance in an introductory university physics course,” *Phys. ST PER* **10**(1), 0101120-1–9 (2014).

²⁰ V.P. Coletta and J.A. Phillips, “Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability,” *Am. J. Phys.* **73**(12), 1172–1182 (2005).

²¹ Turhan Carroll, “A Study of the Effectiveness of Using PER-Based Reforms in a Summer Setting,” (2006), available online at
[<http://www.colorado.edu/physics/Web/reu/Projects/Projects 2006/Turhan Carroll.pdf>](http://www.colorado.edu/physics/Web/reu/Projects/Projects 2006/Turhan Carroll.pdf)
(Retrieved June 26, 2013).

²² R.K. Thornton and D.R. Sokoloff, “Assessing student learning of Newton’s laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula,” *Am. J. Phys.* **66**(4), 338–352 (1998).

²³ L. Hsu, “Measuring the Effectiveness of Summer Intensive Courses for Gifted Students: A Pilot Study and Agenda for Research,” *The Gifted Child Quarterly* **47**(3), 212–218 (2003). As Hsu pointed out in a private communication, not studied is how average students would do in a compressed-format summer course.

²⁴ W.C. Howell, and L.T. Johnson, “An Evaluation of the Compressed-Course Format for Instruction in Accounting,” *The Accounting Review* **62**(2), 403–413 (1982).

²⁵ C. Sheldon, “Student Performance in Compressed Basic Skills Courses,” available online at <http://www.mjc.edu/facultyinformation/research/report_basic_skills_performance_Cerritos_College.pdf> (Retrieved June 22, 2013).

²⁶ C.Q. Sheldon and N.R. Durdella, “Success Rates for Students Taking Compressed and Regular Length Courses in the Community College,” *Community College Journal of Research and Practice* **34**(1–2), 39–54 (2010).

²⁷ C. Hoellwarth and M.J. Moelter, “The implications of a robust curriculum in introductory mechanics,” *Am. J. Phys.* **79**(5), 540–545 (2011).