The relationship between mathematics preparation and conceptual learning gains in physics: a possible "hidden variable" in diagnostic pretest scores

David E. Meltzer
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

Abstract

There have been many investigations into the factors that underlie variations in individual student performance in college physics courses. Numerous studies report a positive correlation between students' mathematical skills and their exam grades in college physics. However, few studies have examined students' learning gain resulting from physics instruction, particularly with regard to qualitative, conceptual understanding. We report on the results of our investigation into some of the factors, including mathematical skill, that might be associated with variations in students' ability to achieve conceptual learning gains in a physics course that employs interactive-engagement methods. It was found that students' normalized learning gains are not significantly correlated with their pretest scores on a physics concept test. In contrast, in three of the four sample populations studied it was found that there is a significant correlation between normalized learning gain and students' pre-instruction mathematics skill. In two of the samples, both males and females independently exhibited the correlation between learning gain and mathematics skill. These results suggest that students' initial level of physics concept knowledge might be largely unrelated to their ability to make learning gains in an interactive-engagement course; students' pre-instruction algebra skills might be associated with their facility at acquiring physics conceptual knowledge in such a course; and between-class differences in normalized learning gain may reflect not only differences in instructional method, but student population differences ("hidden variables") as well.

I. INTRODUCTION

A primary goal of research in physics education is to identify potential and actual obstacles to student learning, and then to address these obstacles in a way that leads to more effective learning. These obstacles include factors that originate during instruction – such as instructional method – as well as those that relate to students' pre-instruction preparation. Previous studies have examined various pre-instruction factors that may or may not be related to students' performance in physics, with mathematics skill being the most common factor. However, in almost all of these studies, the measures of performance adopted were student grades on course exams that emphasized quantitative problem solving. Only in a few cases was students' conceptual knowledge assessed through the use of qualitative problems. And with only a handful of exceptions, there was no attempt to directly measure the gain in student understanding that resulted from instruction.

This paper examines students' mathematics skills and their initial physics conceptual knowledge as factors that may underlie variations in student learning. Learning gain is assessed through pre- and post-testing using a qualitative test of physics conceptual knowledge. One objective of the present study is to determine whether individual students' learning gains are correlated with their initial level of conceptual knowledge as measured by pretest scores on the physics concept test. Another objective is to determine whether those learning gains are correlated with the students' mathematics skills, as determined by pre-instruction testing by a college entrance exam or an algebra/trigonometry skills exam.

In Secs. II and III, I review the results and limitations of previous studies on the relation of students' pre-instruction preparation to their performance in physics courses. In Sec. IV I describe a widely adopted measure of student learning called "normalized learning gain" and explain why it is an appropriate measure for the objectives of this study. In Sec. V various factors that may be related to learning gain are discussed, and the motivation of the present study is presented. The context, methods, and results of the present study are described in Secs. VI, VII, and VIII respectively, and the results are discussed in Sec. IX. The limitations of this study are outlined in Sec. X, and implications for instruction are examined in Sec. XI. The methodological implications of this study for physics education research are addressed in Sec. XII, and Sec. XIII briefly summarizes the main results.

II. PREVIOUS RESEARCH ON THE RELATION OF VARIOUS FACTORS TO STUDENTS' PERFORMANCE IN PHYSICS COURSES

A. Students' mathematical preparation

Many studies appear to show that mathematical ability (mathematical aptitude or accumulated procedural knowledge) is positively correlated to success in traditional introductory physics courses that emphasize quantitative problem solving. Most of these studies have involved college physics students; some have examined the preparation that these students received in high school. Some studies have found a positive correlation between physics course grades and scores on the mathematics part of college entrance exams.[1,2] Many investigators have found positive correlations between grades in college physics and a mathematics skills pretest administered at or near the very beginning of the course. Typically, these pretests involve algebra and trigonometry, although most investigators do not provide samples of their tests.[3-8]

The correlation between mathematics skill and physics performance has not been observed to hold consistently. Reported correlation coefficients vary widely and are not statistically significant for all groups tested. For example, one study found that the overall correlation between grades and an algebra pretest was not significant for males ($r = +0.10$), while for females the correlation was highly significant ($r = +0.48$).[8]

All the studies cited have focused on student performance either on a single physics course exam or on a mean grade from several such exams. In contrast, Hake *et al*.[9] and Thoresen and Gross[10] have reported preliminary investigations of student learning gains in physics courses, determined by both pre-instruction and post-instruction testing. They found that students with the highest learning gains in physics had scored higher on a mathematics skills test than students with the lowest learning gains.

Several investigators have found positive correlations between grades earned by students in their college physics courses and their previous experience and/or grades in either high-school, college mathematics courses, or high-school physics courses.[11,12] However, the overall weight of the literature on factors related to college students' performance in introductory physics is that the measurable impact on performance is substantially larger for mathematics skills as determined by pre-instruction testing, than it is from any measure derived simply from students' experience or lack of it in previous physics or mathematics courses.

B. Students' reasoning skills and other factors

Another factor that has been studied extensively is the possible relation between precourse measures of students' reasoning ability and their college physics grades. Significant correlations between these variables have been reported by numerous investigators[2,4-6,8,13] However, the reported correlations are not significant for all groups, and in most cases the reports do not provide samples of the specific questions used to assess reasoning ability. Recently, Clement[14] has reported a positive correlation between a pretest measure of reasoning ability and learning gain in a high-school physics course.

Other factors that have been found significant to one degree or another are students' achievement expectations,[15] homework grades,[6] high-school GPA,[11,12] college GPA,[16] and a variety of cognitive and emotional factors.[17] A large number of significant preparation and demographic factors were identified by Sadler and Tai.[12] Two studies[4,7] found that students' performance on a pretest of physics conceptual knowledge had a significant positive correlation with course grades.

III. LIMITATIONS OF PREVIOUS RESEARCH

Almost all of the investigations discussed in Sec. II used students' scores (or grades derived from those scores) on physics course exams as a performance measure. It is very likely that in most cases, all or most of the exam questions would be described as traditional quantitative physics problems, although in most cases the nature of the questions was not discussed explicitly. There is by now large body of literature[18-24] that demonstrates convincingly that good performance on such problems does not necessarily indicate good understanding of the physics concepts involved. Performance on such traditional problems may not even be highly *correlated* with conceptual understanding. [24] Our conclusion is that virtually all previously published studies on the relationship between mathematics preparation and physics course performance leave open the question of how, and whether, such preparation may be related to conceptual understanding of physics.

Although various factors – such as mathematics preparation – may be correlated with students' performance on physics exams, this correlation is not direct evidence that there is a causal relationship between the two. To our knowledge, no studies directly test for such a relation. Therefore, it would be improper to conclude from previous studies that, for instance, requiring students to practice and improve their mathematics skills before beginning college physics would necessarily improve their performance in these courses.

Another important limitation of previous research is its failure to examine student learning. A student's performance on a course exam is an indication of the student's knowledge state at the time of the exam, and is not necessarily related to what the student has learned in a particular course. Hence, it is necessary to have some measure of student learning, in contrast to a measure that merely quantifies students' knowledge. One way to provide such a measure is to test students both at the beginning and at (or near) the end of a course to assess how much they may have learned. In this way we can obtain a measure of students' learning gain, which is the quantity that, in principle, is most susceptible to change by actions of the instructor and students during the course. Students' performance on course exams may or may not be correlated with learning gain, and the relationship between performance and learning gain is, at best, an indirect one. Nearly all previous studies have failed to directly investigate the possible relationship of mathematics (and other) preparation to students' learning gain in a college physics course.

IV. NORMALIZED LEARNING GAIN: A KEY MEASURE OF STUDENT LEARNING

The question of how to measure learning gain is not simple and is subject to many methodological difficulties.[25] Because the maximum on a diagnostic instrument is 100%, it is common to observe a strong negative correlation between students' absolute gain scores (posttest minus pretest score) and their pretest scores: higher pretest scores tend to result in smaller absolute gains, all else being equal. For example, in Hake's study of 62 introductory physics courses, absolute gain scores on the Force Concept Inventory (FCI) were significantly (negatively) correlated with pretest score ($r = -0.49$).[20] An alternative is to normalize the gain score to account for the variance in pretest scores. Such a measure is g, the normalized gain, which is the absolute gain divided by the maximum possible gain:

$$g = \frac{posttest\ score - pretest\ score}{maximum\ possible\ score - pretest\ score}.$$

Hake found that $\langle g \rangle$, the mean normalized gain, on the FCI for a given course was almost completely uncorrelated ($r = +0.02$) with the mean pretest score of the students in the course.[26] Therefore, the normalized gain seems to be relatively independent of pretest score. This independence leads us to expect that if a diverse set of classes has a wide range of pretest scores but all other learning conditions are similar, [xx note change xx] the values of normalized learning gain measured in the different classes would not differ significantly. This pretest-independence of the normalized gain also suggests that a measurement of the difference in <*g*> between two classes having very different pretest scores would be reproduced even by a somewhat different test instrument which results in a shifting of pretest scores.

Empirical evidence for this hypothesis is provided by an analysis of the data from Table II of Ref. 21. Students' knowledge of mechanics concepts was tested with two different diagnostic instruments, the FCI, and the Force and Motion Conceptual Evaluation (FMCE).[22] The pretest scores and absolute gain scores yielded by the two instruments were significantly different, but the normalized gains were statistically indistinguishable. The most persuasive empirical support for use of <*g*> as a valid and reliable measure is that <*g*> has now been measured for tens of thousands of students in many hundreds of classes worldwide with extremely consistent results for classes at a broad range of institutions with widely varying student demographic characteristics (including pretest scores).[27]

V. FACTORS THAT MAY BE RELATED TO NORMALIZED LEARNING GAIN

An obvious question is, What are the factors that are related to *g*? Is *g* related to instructional method, or to individual characteristics of the students and their pre-instruction knowledge state?

Hake's original investigation[20] focused on <*g*> for mechanics courses as determined by pre- and post-testing of the FCI. He distinguished two separate groups of courses: (1) those taught with interactive-engagement (IE) methods, and (2) traditional courses that make little or no use of IE methods. Many studies have been published that broadly confirm Hake's major findings,[27] which are that normalized learning gain <*g*> as measured by the FCI in introductory mechanics courses is (1) largely independent of class mean pretest score; (2) virtually independent of the instructor when traditional instructional methods are used; and (3) tends to be significantly higher (by a factor of about two or more) when IE methods are used in comparison with traditional instructional methods. The issue of what *other* factors may be related to variations in *g*, besides instructional method, has, with few exceptions, not been addressed.

Another way of investigating the factors that are related to *g* is to examine the *g* scores of *individual* students to see if the characteristics of individual students may be related to their own learning gains. Hake *et al.*[9] found indications that students' mathematics skills and spatial visualization abilities might be related to their normalized learning gain, and similar results were reported in Ref. 10. Research on high-school students has led Clement to suggest[14] that reasoning ability may be an independent factor. Preliminary data reported in Ref. 28 strongly suggest that there may be a certain amount of variation in $\langle g \rangle$ that can be ascribed to pretest scores (that is, students' initial degree of physics conceptual knowledge). However, in a separate study,[21] the correlation between $\langle g \rangle$ and pretest scores was very low: $r = -0.06$ on FCI; $r = +0.16$ on FMCE.

The objective of the present study is to aid in building a model of the factors that significantly affect students' learning success in physics. To this end, we examine individual students' normalized learning gain scores using a qualitative test of physics conceptual knowledge; students are tested both before and after instruction. We hope to determine (1) whether individual learning gains are correlated with students' initial level of conceptual knowledge as measured by pretest scores on the same physics concept test, and (2) if those learning gains are correlated with the students' mathematics skills, as determined by pre-instruction testing by a college entrance exam or an algebra/trigonometry skills exam.

VI. CONTEXT OF THIS STUDY

This investigation was carried out in the second semester of a two-semester algebra-based general physics sequence. The data reported here originate in four courses taught by the author: two at Southeastern Louisiana University (SLU) in Fall 1997 and Spring 1998, and two courses taught at Iowa State University (ISU) in Fall 1998 and Fall 1999. The number of students in each course ranged from 65 to 92. The focus of the course was electricity and magnetism, including DC circuits. The SLU course consisted of three 50-minute meetings each week held in the lecture room. (A separate lab course was optional and was not taught by the lecture course instructor; there was no recitation session.) At ISU, in addition to three weekly 50-minute meetings in the lecture room, there is one 50-minute recitation session each week. (There is also a separate required lab in which the lecture instructor has only limited involvement.) These courses made much use of IE instructional methods and employed a variant of Mazur's Peer Instruction.[24,29] The primary curricular material was the *Workbook for Introductory Physics.*[29] Instruction in the recitation sessions at ISU was modeled closely on the University of Washington tutorials,[23] although most of the material used came from the *Workbook for Introductory Physics.*

VII. METHODS

Students' conceptual knowledge was assessed by the administration of a physics concept diagnostic test on the first and last days of class; only students who took both pre- and posttests are part of the sample. Students' preinstruction mathematics skill was assessed by their score either on the ACT Mathematics Test or on an algebra-trigonometry skills test. A variety of statistical tests were then performed to assess the relation (if any) between students' individual normalized learning gain, and their pre-instruction scores on both the physics concept test and the mathematics skills test.

The diagnostic instrument was the Conceptual Survey in Electricity (CSE). This 33-item multiple-choice test surveys knowledge related to electrical fields and forces and the behavior of charged particles. The questions on the CSE are almost entirely qualitative. About half of the items are also included on the Conceptual Survey in Electricity and Magnetism (CSEM).[19] The creators of the CSEM remark that it contains "a combination of questions probing students' alternative conceptions and questions that are more realistically described as measuring students' knowledge of aspects of the formalism."[19]

On the pretest, students were given enough time to respond to all 33 questions. Neither grades nor answers for this pretest were posted or discussed. On the last day of class, the same CSE was administered as an extra-long in-class quiz. However, students were asked to respond to only 23 of the questions.[30] The CSE was used in this abridged form for various reasons. For example, in some cases, the notational conventions differed from what was used in class (for instance, electric field lines are used on the CSE, but only field vectors were used in class). In other cases, the questions involved material that was covered peripherally or not at all in class. Only the 23 designated items were graded, both on the pretest and the posttest. All CSE scores discussed in this paper (as well as quantities derived from them) refer only to the 23-item abridged CSE.

For the SLU samples, scores on the ACT Mathematics Test were used to assess pre-instruction mathematics skill. This test is a college entrance exam, and so there is typically a 1-3 year gap between the time students take this test and the time they take the CSE. The instrument used at ISU is a 38-item multiple-choice test originally developed by Hudson during the course of his investigations (cited in Sec. II) into the effect of mathematics preparation on students' physics performance. It includes the following topics among others: solving and manipulating one- and two-variable algebraic equations; factoring quadratic equations; unit conversions; elementary trigonometry; straight-line graphs; powers-of-ten notation; simple word problems; and addition of numerical and algebraic fractional expressions. (See Appendix A for representative problems.)

All students who register for the first semester course in the algebra-based physics sequence at ISU are required to take this test; it does not count toward the students' grade. Because students take this exam at the beginning of the *first* semester course, there was a gap of at least two months (as in the case of summer-school students) between when they took the mathematics test and when they took the CSE. More often, the gap was five to 12 months.

Several modifications were introduced during the ISU 1999 course which, it was hoped, would improve instruction. Both graduate student teaching assistants for the course were members of the Physics Education Research Group and had extensive experience and capabilities in inquiry-based instruction. For many of the recitation-session/tutorials, an additional undergraduate teaching assistant was present. During this course, both the teaching assistants and the course instructor spent many out-of-class hours in individual instruction with students who solicited assistance.

VIII. RESULTS

A. CSE Pretest scores are not correlated with individual normalized learning gain

Table I shows the correlation coefficients between individual students' *g* scores and their CSE pretest score for the four samples. The correlations are very small and none is close to being statistically significant. Figure 1 shows the value of *g* and the CSE pretest score for all students in the ISU 1998 sample. The correlation coefficient for this relation is $r = 0.00$; there is no evidence of any pattern in the data points. This random pattern is typical of all four samples.

Table II presents comparisons of $\langle g \rangle$ for several different subgroups of two different samples.[31] For the 1998 sample in Table II, "Top half" refers to the students with the 29 highest scores on the CSE pretest; "Bottom half" refers to the group with the 30 lowest CSE pretest scores. (The 59-student sample was divided in this way to form two groups of nearly equal size; the groups had zero overlap in pretest scores. Pretest scores ranking #24-29 were identical [eight correct], and scores in the group #30-43 were equal [seven correct]). This method was used to form the other subgroups represented in Tables II and IV.) The mean CSE pretest scores of these two groups were very different, but their normalized gains were not statistically distinguishable according to the one-tailed *t*-test: $\langle g_{top\ half} \rangle = 0.68$, $\langle g_{bottom\ half} \rangle = 0.63$, $t = 0.84$, $p = 0.20$. A comparison between even more disparate groups is also shown in Table II. "Top quartile" refers to students with the 15 highest CSE pretest scores in the 1998 sample, while "Bottom quartile" refers to the 16 lowest in that sample. The normalized gains of these two groups were virtually identical. Table II also presents a similar set of comparisons for the ISU 1999 sample. The results for this sample share the main characteristic of the 1998 sample, even for the extreme "Top fifth" and "Bottom fifth" groups: $\langle g_{top\ fifth} \rangle = 0.73$, $\langle g_{bottom\ fifth} \rangle = 0.67$; these gains are not significantly different according to the one-tailed t-test ($t = 0.98$, $p = 0.17$).

Figure 2 shows the distributions of the normalized gain among the Top half and Bottom half groups from the 1998 sample; there are no striking differences between the pretest groups. A similar result was found for the 1999 sample. This result reinforces the conclusion from the correlation analysis that the pretest score on the CSE is not a significant factor in determining a student's normalized learning gain.

B. Mathematics pretest scores are correlated with normalized learning gain

Table III presents the correlation coefficient and corresponding statistical significance (that is, p value) for the relation between students' g scores and their scores on the pre-instruction mathematics skills test. The correlation for the SLU 1998 sample was not statistically significant; the correlations for the other three samples were all statistically significant at the $p < 0.01$ level.

Figure 3 shows g as a function of score on the Mathematics Diagnostic Test for the ISU 1998 sample. A positive correlation between the two variables is evident. A similar correlation (though not as large) is also evident in the SLU 1997 and ISU 1999 sample data. Examination of the residuals, that is, the differences between data points and regression fit line, shows that there are no marked nonlinearities evident in the data, and further that the sample variances are fairly uniformly distributed (that is, the data are "homoscedastic").

Table IV presents comparison data for subgroups chosen in a manner analogous to that used in Table II. For instance, the first two lines compare $<g>$ for the group of students in the ISU 1998 sample with the highest math pretest scores (Top half, actually the top 47%) to the group with the lowest scores in the same sample (Bottom half, the lowest 53%). In this case – in sharp contrast to the situation in Table II – the learning gains of the two groups are very different, with high statistical significance: $<g_{top\ half}> = 0.75$, $<g_{bottom\ half}> = 0.56$; $p = 0.0001$ (one-tailed). When we go to groups even further separated by their mathematics pretest scores – the top quartile and bottom quartile groups – we find an even greater difference between their mean normalized gain: $<g_{top\ quartile}> = 0.77$, $<g_{bottom\ quartile}> = 0.49$, $p = 0.001$ (one-tailed).

Also shown in Table IV is an analogous set of data for the ISU 1999 sample. The differences in $<g>$ between the Top half and Bottom half mathematics pretest groups are substantially smaller than in the 1998 sample, but are still statistically significant: $<g_{top\ half}> = 0.75$, $<g_{bottom\ half}> = 0.66$, $p = 0.04$ (one-tailed). Moreover, the difference in learning gain is substantially larger for the groups closer to the extremes of the mathematics pretest score range, that is, the Top quartile and Bottom quartile groups: $<g_{top\ quartile}> = 0.78$, $<g_{bottom\ quartile}> = 0.60$, $p = 0.005$ (one-tailed). This difference is consistent with the data from the 1998 sample and significantly strengthens the case that the observed correlation is real and not an artifact produced by the particular selection of the subgroups.

Figure 4 shows the population distributions for the normalized gain for the ISU 1998 sample, portraying the top and bottom mathematics pretest score groups. There is a very noticeable skewing of the distributions toward the high end of the *g* scale for the high math group. Again, this result is consistent with the correlation analysis and is in striking contrast to the distributions shown in Fig. 2.

It is worth noting another feature of Table IV. Although the normalized gains for the Top half and Top quartile groups in the 1999 sample are nearly identical to those for the corresponding groups in the 1998 sample, that is not the case for the Bottom half and Bottom quartile groups. The *g*'s for those groups are substantially larger in the 1999 sample. It is tempting to ascribe these higher *g* values to the differences in the instructional methods implemented in 1999, although this is merely speculation.

C. The math score/learning gain correlation is present for both males and females

Table V presents the correlation coefficients and corresponding statistical significance for the male and female subgroups of the two ISU samples (selected because they are larger and contain more reliable data). Although the value of r for males in the ISU 1998 sample is larger than that for females, the difference is not statistically significant ($p = 0.50$, using Fisher transformed values[32]). In the 1999 sample, the correlation coefficients for males and females are nearly identical. All four correlations are statistically significant at the $p < 0.05$ level for a one-tailed test, warranted in this case given the positive correlation observed for both full samples.

IX. DISCUSSION OF RESULTS

The results in this study regarding the *lack* of correlation between normalized learning gain and CSE pretest score are very consistent. However, the results for the mathematics pretest score are in striking contrast to those for the CSE pretest score: in three of the four samples, there is a significant positive correlation ($p < 0.01$) between normalized learning gain and mathematics pretest score. This relation observed between normalized learning gain and pre-instruction mathematics skill is consistent with the preliminary results presented in Refs. 9 and 10; however, the present study represents the first comprehensive examination of this relation.

Another way to look at the data is to compare the mathematics pretest scores for high gainers and low gainers. Hake *et al.*[9] arbitrarily define high and low gainers as those with $g \geq 1.3\langle g\rangle$ and $g \leq 0.7\langle g\rangle$, respectively, where $\langle g\rangle$ is the mean for the class. They found that high gainers scored 19% higher on the mathematics skills pretest than did the low gainers in their sample. If we apply their definitions and examine mean mathematics pretest scores $\langle m\rangle$ (m is the percentage of correct responses), we find that $\langle m\rangle_{\text{high gainers}} = 81\%$, $\langle m\rangle_{\text{low gainers}} = 60\%$ for ISU 1998 and $\langle m\rangle_{\text{high gainers}} = 80\%$, $\langle m\rangle_{\text{low gainers}} = 65\%$ for ISU 1999. These results are remarkably consistent with those reported in Ref. 9.

The results of Ref. 8 suggested that any observed correlation might *not* be a general characteristic of all students, but of females only. Just as CSE pretest scores were a potentially confounding variable, students' gender has to be considered one as well. With this consideration in mind, the fact that results for *both* ISU samples show statistically indistinguishable correlation coefficients for male and female subpopulations is very significant. Moreover, all four of these correlations were significant at the $p < 0.05$ level (one-tailed test) for their individual subpopulation.

The relatively low correlation coefficients found in this study (between +0.30 and +0.46) yield little predictive power regarding the expected value of the learning gain of an *individual* student, based on his or her pre-instruction score on the mathematics skills test. On the other hand, when assessing the likelihood of a student becoming a high gainer or a low gainer (defined, in this case, as one with gains above or below the class median, respectively), considerably more predictive power is possible. For instance, if we look at the students in the ISU 1998 sample with the lowest mathematics scores (the Bottom quartile in Table IV), we find that only 21% of them (3 of 14) have gains above the class median of $g = 0.693$. In comparison, among the group with the highest mathematics scores (Top quartile), 77% (10 of 13) have gains above the class median. Therefore, knowledge of whether a student had unusually high or low mathematics scores could have allowed a fairly high-confidence prediction of whether they would end up with above- or below-average gains.

In striking contrast to this predictability based on mathematics pretest score, the knowledge of a student's CSE pretest score would have allowed no such prediction. The group with the lowest CSE pretest scores (Bottom quartile in Table II) had 50% (8 of 16) with gains above the class median. At the same time, the group with the highest CSE pretest scores (Top quartile in Table II) also had the same number of above-median and below-median gains (7 of each, with one student at exactly the class median).

Higher predictive power is associated with the mean learning gains of the subgroups at the high and low ends of the mathematics scale. The students in the ISU 1998 sample with the lowest mathematics scores have an expected normalized gain (95% confidence interval) ranging from 0.35 to 0.64. In comparison, the expected gain of the group with the highest scores on the mathematics exam range from 0.68 to 0.85. Therefore, we can be highly confident that – for an equivalent sample – the mean gain of the lowest mathematics group would be *below* the class mean of 0.65, while that of the highest mathematics group would be *above* the mean. Obviously, no comparable statement could be made about the groups with the lowest and highest CSE pretest scores. The correlations observed for the other samples are lower, and therefore so is the predictive power, but the same pattern persists.

X. LIMITATIONS OF THIS STUDY

A. *Student population.* Students enrolled in calculus-based physics courses often have a much more substantial mathematics background than those in the algebra-based course used in this study; this background may be associated with a different relation between mathematics skills and conceptual learning gain in physics. It should also be noted that the population of the two ISU samples was 60% female, a high proportion in comparison to the calculus-based course.

B. *Subject Matter.* Students have considerably less day-to-day experience and accumulated common sense notions regarding electric and magnetic phenomena in comparison with mechanics. Many of the concepts studied (for example, the electromagnetic field) are considerably more abstract than most encountered in the introductory mechanics course. It is conceivable that if a comparable study were done in connection with student learning in a less abstract and more familiar domain, and if assessment relied less on interpretation and analysis of formal representations, the results might be different.

C. *Instructional Methods.* The instructional methods used in this study were certainly not comparable to traditional methods of instruction in widespread national use. They made much use of IE methods, including interactive lecture[29] and group work in the style of the University of Washington tutorials. On the exams, quizzes, and homework, the emphasis was very much on the type of qualitative questions that are used on the Conceptual Survey in Electricity (without teaching to the test). Overall normalized gains were unusually high by national standards. It is possible that the results reported in this study are related in some fashion to the courses' instructional emphasis on qualitative and conceptual problem solving.

D. *Hidden variables.* It is an inherent limitation of any study that relevant variables might be neglected. For a study such as this one, the particular danger is that some of the neglected variables might actually be so important that their omission is ultimately the source of a spurious apparent correlation that would disappear if these variables had been included. This can happen if the neglected variable is strongly correlated with the targeted dependent variable (learning gain, in this case.)

For example, logical reasoning ability is a variable that some investigators have found to be significant. Suppose that logical reasoning ability is strongly correlated with physics learning gain, and moreover that this reasoning ability is also strongly correlated with pre-instruction mathematics skill. We might find that, for a given level of reasoning ability, there is no separate correlation between mathematics skill and physics learning gain. That would imply that improving reasoning ability might improve learning gain, but that improving mathematics skill would not have such an effect in the absence of any accompanying changes in reasoning ability.

XI. IMPLICATIONS FOR INSTRUCTION

The evidence from this study is that in an IE course, students' normalized learning gains on the CSE are essentially *independent* of their pretest scores. The implication is that, at least with this type of instruction, students' potential to achieve gains in understanding is independent of whether they begin the course with high, low, or even zero initial levels of physics concept knowledge. Knowledge of students' CSE pretest scores might allow some prediction of their probable final level of understanding, but would allow no prediction of their ultimate learning gains. This result is encouraging because it implies that students have an equal chance at learning regardless of their initial knowledge of concepts in electricity.

Although students' initial level of physics concept knowledge may have no impact on their learning gains, the same cannot be said for their initial level of mathematics skill. In three of the four samples in this study, students with higher levels of preinstruction mathematics skill had substantially higher learning gains on the physics concepts – independent of their initial knowledge of those concepts – when compared to students with lower mathematics skill levels (true for both males and females at ISU).

Whether or not this correlation would hold up if other variables, unknown and therefore hidden to us, were included in the analysis is irrelevant to the potential utility of mathematics skill as an indicator of probable high- and low-gainers. If there are indeed other relevant variables associated with learning gain, it seems likely that they would be correlated with mathematics skill. Until they are known, mathematics skill may be used as a substitute measure for those variables – perhaps not so directly related as those other (hypothetical) variables to the targeted parameter of learning gain, but associated with it nonetheless. (The possibility of using mathematics skill as an indicator of physics learning potential was suggested in Ref. 9 and by many of the investigators cited in Sec. II.) It should be emphasized that the correlation observed between mathematics preparation and normalized learning gain does not imply that mathematics skill is *causally* related to physics concept learning gains. It simply means that whatever factors may ultimately be found to be causally related to learning gain, mathematics skill is probably associated with them in some manner.

In the same sense in which the lack of g versus CSE pretest score correlation was encouraging, the positive correlation between g and a mathematics pretest score is somewhat disconcerting. The implication may be that students with lower levels of pre-instruction mathematics skills (whatever the cause) may be unlikely as a group to attain a level of physics learning gain achieved by those with greater mathematics skill, all else being equal. An instructor who transports instructional methods and curricula from one student population to another with much lower mathematics skill levels might find that lower learning gains are achieved. However, the poorer expected outcome of using the *same* instruction with students of lower mathematics skill leaves open the possibility that different instructional methods and curricula might ultimately achieve the same levels of learning gain success with the new population as with the old. The higher learning gains of the low-math group in the ISU 1999 sample (which received modified instruction) might offer some mild support for this speculation.

XII. METHODOLOGICAL IMPLICATIONS

A. The observed correlations might imply that widely diverse populations taught with identical instructional methods might manifest different normalized learning gains.

The low-math and high-math subgroups in this study were taught with identical instructional methods (for all practical purposes). And yet it is clear that their mean normalized learning gains were significantly different. If one imagines an entire class populated with low-math students at institution A, and a different class – perhaps at a different institution B – populated with high-math students, it is plausible that instruction carried out with identical methods and materials – perhaps with the identical instructor – might nonetheless result in different values of $\langle g \rangle$ for the two classes.

The extent of the variation in g in a given population that might be ascribed to variations in mathematics preparation would depend on the range of mathematics skills represented in that population; it could be estimated by using the linear regression equation that is a best fit to the g versus M_{pre} data, where M_{pre} is the mathematics pretest score (for example, the data shown in Fig. 3). Using this method, we estimate for the ISU samples that variations in $\langle g \rangle$ ascribable solely to the average variability of students' mathematics preparation (that is, for students having M_{pre} within the range $\langle M_{pre} \rangle \pm 1.0$

s.d., where *s.d.* is the standard deviation of the M_{pre} scores) are confined to the range <*g*> ≈ <*g*>$_{\text{mean}}$ ± 0.15<*g*>$_{\text{mean}}$.

If we speculate that mechanics courses would show correlations between normalized gain and mathematics preparation similar to those in this study, we can estimate that the variation in <*g*> ascribable to mathematics preparation would be ± 0.07 for <*g*> ≈ 0.45 (a typical value for mechanics courses that employ interactive engagement). This variation is much smaller than the difference commonly found between courses taught with IE and traditional methods, respectively.

B. It may be necessary to consider possible second-order effects due to sample-to-sample differences in pre-instruction knowledge state.

This particular statement can easily be put in a familiar context. The author measured <*g*> on the CSE to be ≈ 0.48 in his courses at SLU. After attempting to improve his instructional methods and materials, he found <*g*> ≈ 0.67 in the courses he taught at ISU. (Mean CSE pretest scores were 28% at SLU, 32% at ISU.) Does this difference imply that he succeeded in improving his instruction? Does the large apparent gain in <*g*> perhaps overstate the actual improvement? This type of practical question is one that we often attempt to answer with pre-/post-test data.

If one is actually planning an experiment in which <*g*> is to be a measure of comparative learning gains, it is standard practice to randomize the different samples so that the effects of any potential uncontrolled variables (such as mathematics preparation) may be expected to cancel each other out. One can argue that <*g*> should never be used to compare potentially non-equivalent (that is, non-randomized) samples. The author's courses at SLU and ISU are a good example of this problem. Should one directly compare the <*g*>'s in the two cases, or is some set of hidden variables at work, variables that actually make the two student samples not equivalent?

It is important to emphasize that there is no reason to believe that effects of hidden variables – even combined – are likely to be of the same scale as the two-standard-deviation differences in <*g*> on the FCI between traditional instruction and IE instruction documented by Hake. Moreover, with a sample as large as Hake's, it is very unlikely that the IE/non-IE differences in <*g*> could possibly be due to the effects of hidden variables that have not been averaged out. However, when one has much smaller samples in just a few courses taught at widely disparate institutions where the differences in <*g*> may not be so large, there is much more uncertainty in the comparison. To first-order, large differences in <*g*> are probably due to instructional method. However, almost certainly, higher-order effects of unknown scale and origin influence comparative <*g*> statistics in as yet unknown ways.

XIII. SUMMARY

The results of this study provide substantial evidence that factors other than instructional method play a role in determining students' normalized learning gains. Further research to identify and measure these factors should aid in understanding and addressing students' learning difficulties in physics, as well as in analyzing data that result from assessments of student learning.

ACKNOWLEDGMENT

I am very grateful to F. C. Peterson for bringing to my attention the existence of the ISU Mathematics Diagnostic Test data and for many helpful conversations.

Table I. Correlation between normalized learning gain and pretest score on CSE.

Sample	N	Correlation coefficient between student learning gain g and CSE pretest score	Statistical significance (two-tailed)
SLU 1997	45	+0.15	$p = 0.35$ *(not significant)*
SLU 1998	37	+0.10	$p = 0.55$ *(not significant)*
ISU 1998	59	0.00	$p = 0.98$ *(not significant)*
ISU 1999	78	+0.10	$p = 0.39$ *(not significant)*

Table II. ISU samples: Gain comparison, students with high and low CSE pretest scores. <*g*> represents the mean of individual students' normalized gains; s.d.≡ standard deviation.

	N	Mean CSE Pretest Score	<*g*> *(s.d.)*
1998			
Top half	29	44%	0.68 *(0.19)*
Bottom half	30	25%	0.63 *(0.23)*
Top quartile	15	50%	0.65 *(0.21)*
Bottom quartile	16	20%	0.66 *(0.24)*
1999			
Top third	30	43%	0.74 *(0.18)*
Bottom third	27	18%	0.72 *(0.17)*
Top fifth	14	49%	0.73 *(0.20)*
Bottom fifth	15	14%	0.67 *(0.13)*

Table III. Correlation between normalized learning gain and mathematics pretest score.

Sample	N	Correlation coefficient between student learning gain g and mathematics pretest score	Statistical significance (two-tailed)
SLU 1997	45	+0.38	$p < 0.01$
SLU 1998	37	+0.10	$p = 0.55$ *(not significant)*
ISU 1998	59	+0.46	$p = 0.0002$
ISU 1999	78	+0.30	$p < 0.01$

Table IV. ISU samples: Gain comparison, students with high and low mathematics pretest scores. <*g*> represents the mean of individual students' normalized gains. s.d. ≡ standard deviation.

	N	Mean Mathematics Pretest Score	<*g*> *(s.d.)*
1998			
Top half	28	89%	0.75 *(0.15)*
Bottom half	31	63%	0.56 *(0.22)*
Top quartile	13	93%	0.77 *(0.14)*
Bottom quartile	14	49%	0.49 *(0.25)*
1999			
Top half	37	86%	0.75 *(0.20)*
Bottom half	36	55%	0.66 *(0.22)*
Top quartile	21	90%	0.78 *(0.17)*
Bottom quartile	20	44%	0.60 *(0.23)*

Table V. Correlation between normalized learning gain and mathematics pretest score for males and females (ISU samples).

	N	Correlation coefficient between student learning gain *g* and mathematics pretest score	Statistical significance (one-tailed test)
ISU 1998: males	22	+0.58	$p < 0.01$
ISU 1998: females	37	+0.44	$p < 0.01$
ISU 1999: males	33	+0.29	$p = 0.04$
ISU 1999: females	45	+0.31	$p = 0.03$

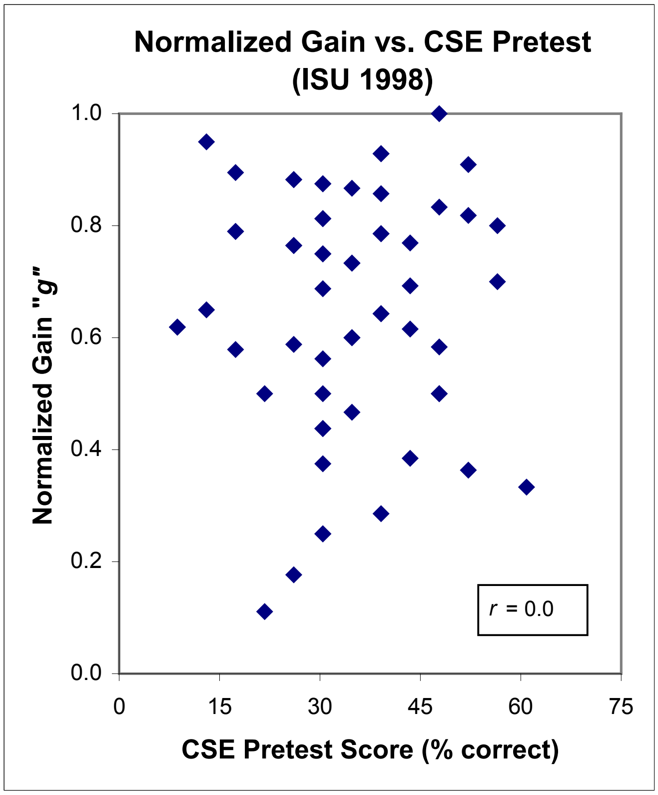
Figure Captions

Figure 1. Scatter plot of ISU 1998 sample; data points correspond to individual students, plotted according to their individual normalized learning gain g score on the Conceptual Survey in Electricity (CSE) and their pretest score on that same exam. Correlation coefficient $r = 0.00$.

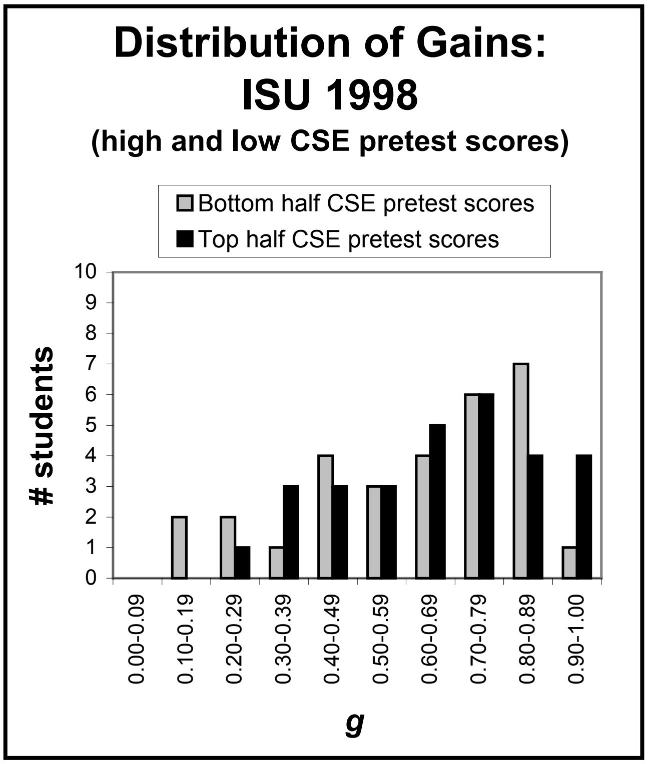
Figure 2. Distribution of normalized learning gains for ISU 1998 sample: light bars, students with 30 lowest scores on CSE pretest ($\langle g \rangle = 0.63$); dark bars, students with 29 highest scores on CSE pretest ($\langle g \rangle = 0.68$). ($\langle g \rangle$ represents the mean of individual students' normalized gains.)

Figure 3. Scatter plot of ISU 1998 sample. Data points correspond to individual students, plotted according to their individual normalized learning gain g on the CSE and their pre-instruction score on the Mathematics Diagnostic Test. Correlation coefficient $r = +0.46$, $p = 0.0002$; the data are best fit by the linear relation $g = 0.228 + 0.01496M$, where M is the number of correct answers on the Mathematics Diagnostic Test (maximum = 38).

Figure 4. Distribution of normalized learning gains for ISU 1998 sample: light bars, students with 31 lowest scores on the Mathematics Diagnostic Test ($\langle g \rangle = 0.56$); dark bars, students with 28 highest scores on the Mathematics Diagnostic Test ($\langle g \rangle = 0.75$).



ms 15834
Figure 1
Meltzer
AJP



ms 15834,
Figure 2
Meltzer
AJP

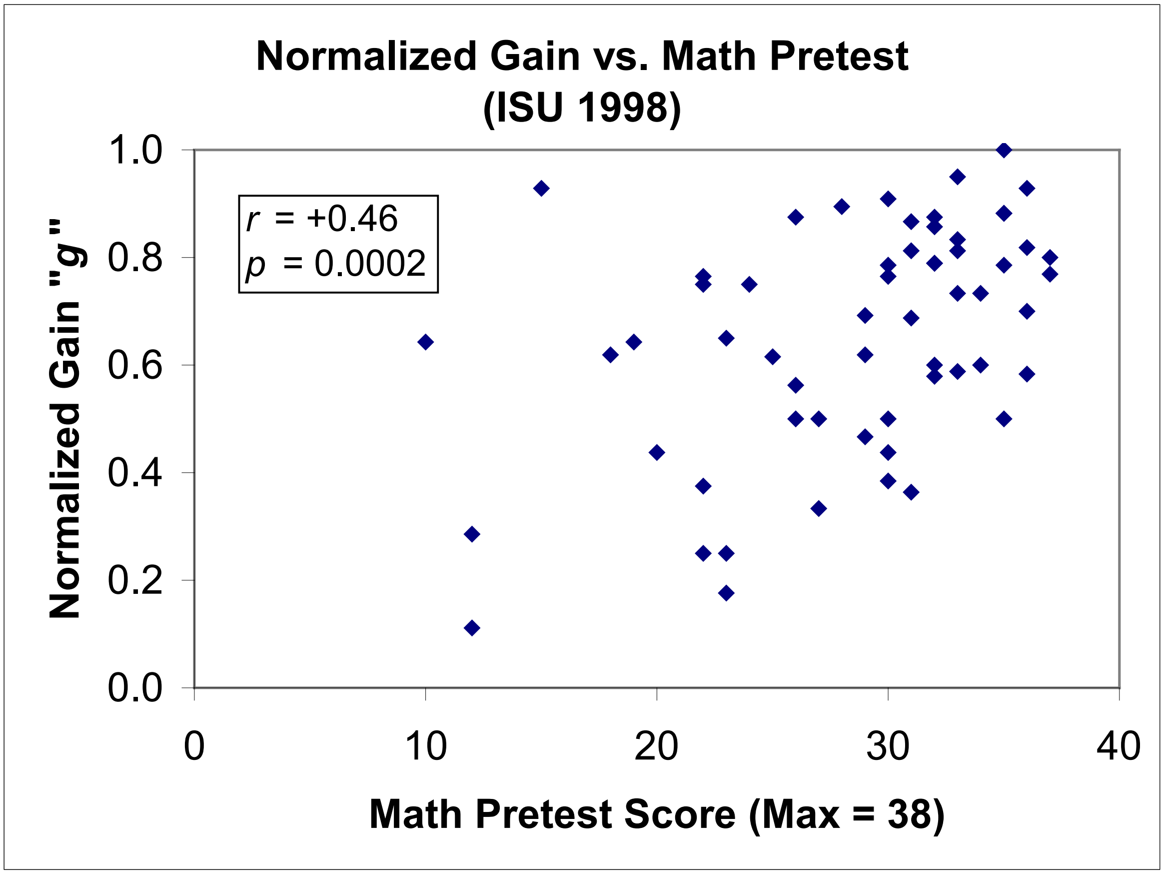
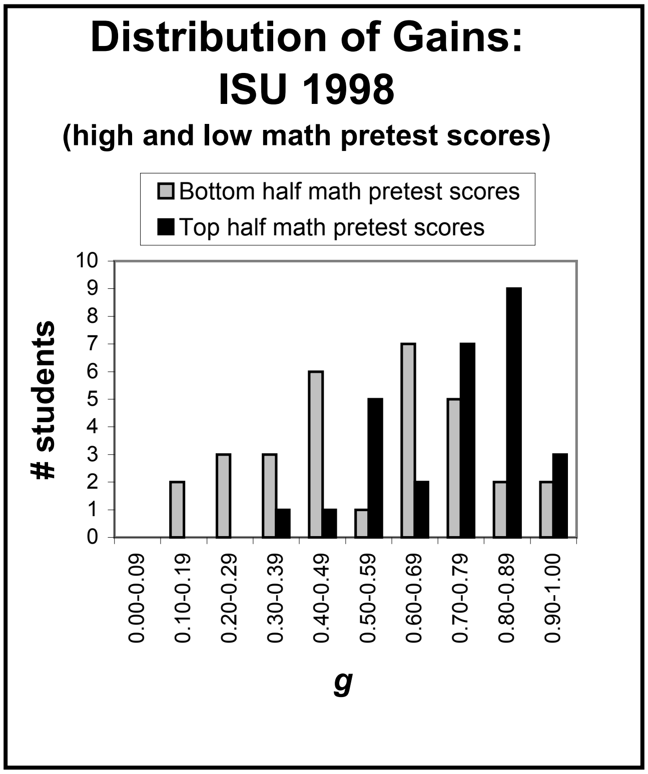

Normalized Gain vs. Math Pretest
(ISU 1998)
r = +0.46
p = 0.0002
Normalized Gain "g"
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Math Pretest Score (Max = 38)

Figure 3



ms 15834,
Figure 4
Meltzer
AJP

Appendix

Selected problems from the Mathematics Diagnostic Test used at ISU (author: H.T. Hudson):

1. $\sqrt{15^2 - 9^2} = ?$

a. 6 b. $\sqrt{6}$ c. 12 d. $\sqrt{12}$ e. $\sqrt{135}$

2. Find y as a function of x from the following equations.

$2x - t = 2$

$y - 4 = 3t$

a. $y = 3x + 4$
b. $y = 10 - 3x$
c. $y = 3x + 6$
d. $y = 4 - 6x$
e. $y = 6x - 2$

3. $\frac{3}{14} + \frac{7}{6} =$ ________.

a. 29/21
b. 21/20
c. 10/21
d. 18/49
e. 5/21

4. If the angle A = 4π/6 radians, what is the value of A in degrees?

a. 60° b. 120° c. 90° d. 45° e. 210°

5. $\frac{12 \times 10^8}{2 \times 10^{-2}} =$ _______.

a. 6×10^{-4} b. 10×10^{10} c. 10×10^{-10} d. 6×10^{10} e. 10×10^{6}

[1]Ralph H. Blumenthal, "Multiple instruction and other factors related to achievement in college physics," Sci. Ed. **45**, 336-342 (1961).

[2]H. Daniel Cohen, Donald F. Hillman, and Russell M. Agne, "Cognitive level and college physics achievement," Am. J. Phys. **46**, 1026-1029 (1978).

[3]H. T. Hudson and W. R. McIntire, "Correlation between mathematical skills and success in physics," Am. J. Phys. **45**, 470-471 (1977); H. T. Hudson and Ray M. Rottmann, "Correlation between performance in physics and prior mathematics knowledge," J. Res. Sci. Teach. **18**, 291-294 (1981); Warren Wollman and Frances Lawrenz, "Identifying potential 'dropouts' from college physics classes," J. Res. Sci. Teach. **21**, 385-390 (1984); H. T. Hudson, "A comparison of cognitive skills between completes and dropouts in a college physics course," J. Res. Sci. Teach. **23**, 41-50 (1986); C. J. Linder and H. T. Hudson, "A comparison of mathematics backgrounds between American and South African physics students," Sci. Ed. **73**, 459-465 (1989).

[4]Audrey B. Champagne, Leopold. E. Klopfer, and John H. Anderson, "Factors influencing the learning of classical mechanics," Am. J. Phys. **48**, 1074-1079 (1980); Audrey B. Champagne and Leopold E. Klopfer, "A causal model of students' achievement in a college physics course," J. Res. Sci. Teach. **19**, 299-309 (1982).

[5]H. T. Hudson and Dov Liberman, "The combined effect of mathematics skills and formal operational reasoning on student performance in the general physics course," Am. J. Phys. **50**, 1117-1119 (1982).

[6]W. T. Griffith, "Factors affecting performance in introductory physics courses," Am. J. Phys. **53**, 839-842 (1985).

[7]Ibrahim Abou Halloun and David Hestenes, "The initial knowledge state of college physics students," Am. J. Phys. **53**, 1043-1055 (1985).

[8]Susan McCammon, Jeannie Golden, and Karl K. Wuensch, "Predicting course performance in freshman and sophomore physics courses: women are more predictable than men," J. Res. Sci. Teach. **25**, 501-510 (1988).

[9]R. R. Hake, R. Wakeland, A. Bhattacharrya, and R. Sirochman, "Assessment of individual student performance in an introductory mechanics course," AAPT Announcer **24** (4), 76 (1994).

[10]Michael J. Thoresen and Carl M. Gross, "Factors contributing to success in college physics," AAPT Announcer **30** (4), 128 (2000).

[11]John R. Bolte, "Background factors and success in college physics," J. Res. Sci. Teach. **4**, 74-78 (1966); Gerald E. Hart and Paul D. Cottle, "Academic backgrounds and achievement in college physics," Phys. Teach. **31**, 470-475 (1993); Brian J. Alters, "Counseling physics students: A research basis," Phys. Teach. **33**, 413-415 (1995).

[12]Philip M. Sadler and Robert H. Tai, "Success in introductory college physics: The role of high school preparation," Sci. Ed. **85,** 111-136 (2001).

[13]George Barnes, "Scores on a Piaget-type questionnaire versus semester grades for lower-division college physics students," Am. J. Phys. **45**, 841-847 (1977); Dov Liberman and H. T. Hudson, "Correlation between logical abilities and success in physics," Am. J. Phys. **47**, 784-786 (1979); Morris A. Enyeart, Dale Baker, and Dave Vanharlingen, "Correlation of inductive and deductive logical reasoning to college physics achievement," J. Res. Sci. Teach. **17**, 263-267 (1980); J. Cilliers, P. Kruger, I. Basson, and P. A. Kirschner,

"Psychometric testing as a predictor of student performance in first year physics practicals," J. Sci. Educ. Tech. **5**, 47-57 (1996).

[14]John M. Clement, "Using physics to raise students' thinking skills," AAPT Announcer **30** (4), 81 (2000); "Techniques used to raise thinking skills in high school," **30** (4), 92 (2000); "The correlation between scientific thinking skill and gain in physics conceptual understanding," **31**(2), 82 (2001).

[15]O. J. Ehindero, "Correlates of physics achievement: The role of gender and non-induced student expectations," J. Expt. Educ. **54**, 189-92 (1986).

[16]M. J. Marr, E. W. Thomas, M. R. Benne, A. Thomas, and R. M. Hume, "Development of instructional systems for teaching an electricity and magnetism course for engineers," Am. J. Phys. **67**, 789-802 (1999).

[17]Christoph von Rhoneck, Karl Grob, Gerhard W. Schnaitmann, and Bruno Volker, "Learning in basic electricity: How do motivation, cognitive and classroom climate factors influence achievement in physics," Int. J. Sci. Ed. **20**, 551-565 (1998).

[18]Lillian C. McDermott and Edward F. Redish, "Resource Letter: PER1: Physics Education Research," Am. J. Phys. **67,** 755-767 (1999) and <http://www.physics.umd.edu/rgroups/ripe>.

[19]David P. Maloney, Thomas L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen, "Surveying students' conceptual knowledge of electricity and magnetism," Phys. Educ. Res., Am. J. Phys. Suppl. **69**, S12-S23 (2001).

[20]Richard R. Hake, "Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66,** 64-74 (1998); <http://www.physics.indiana.edu/~sdi/>.

[21]Karen Cummings, Jeffrey Marx, Ronald Thornton and Dennis Kuhl, "Evaluating innovations in studio physics," Phys. Educ. Res., Am. J. Phys. Suppl. **67**, S38-S44 (1999).

[22]Ronald K. Thornton and David 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**, 338-352 (1998).

[23]Lillian C. McDermott, "Millikan Lecture 1990: What we teach and what is learned – Closing the gap," Am. J. Phys. **59**, 301-315 (1991); Lillian C. McDermott, "Bridging the gap between teaching and learning: The role of research," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399,** 139-165 (1997)], Pt. 1.

[24]Eric Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997), pp. 3-7, and *passim*.

[25]Walter R. Borg and Meredith D. Gall, *Educational Research*, *An Introduction* (Longman, New York, 1989), 5th ed., pp. 728-733.

[26]<*g*> is found by using the mean pretest score and mean posttest score of all students in the class. Note that the notation to be adopted in the present paper is unorthodox; see Ref. 31.

[27]Edward F. Redish and Richard N. Steinberg, “Teaching physics: figuring out what works,” Phys. Today **52**, 24-30 (1999); Richard R. Hake, "Lessons from the physics education reform effort," Conservation Ecology **5**, 28 (2002), <http://www.consecol.org/vol5/iss2/art28>.

[28]Charles Henderson, Ken Heller and Pat Heller, “Common concerns about the Force Concept Inventory,” AAPT Announcer **29**(4), 99 (1999) and <http://www.physics.umn.edu>.

[29]David E. Meltzer and Kandiah Manivannan, “Promoting interactivity in physics lecture classes,” Phys. Teach. **34**, 72-78 (1996); “Transforming the lecture-hall environment: The fully interactive physics lecture,” Am. J. Phys. **70**, 639-654 (2002).

[30]The following item numbers on CSE Form D were used in all cases [more recent versions of the CSE use different numbering; corresponding CSEM item numbers are shown in brackets]: 3, 4, 5, 6, 7, 8, 9[7], 10[8], 11[9], 12, 13[10], 14[11], 15, 16[12], 23[17], 24[18], 25[19], 26, 27[16], 28[20], 31[3], 32[4], 33[5]. In several cases, there are minor differences between the CSE questions and the corresponding CSEM items.

[31]Here the notation <*g*> is used to symbolize the mean value of the *individual* student *g*’s. Its more common use in the literature is to symbolize the *g* calculated by using mean values of the pretest and posttest scores of all students in a class.

[32]David J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures* (Chapman & Hall/CRC, Boca Raton, 2000), 2nd ed., pp. 782-784.