Tale of two curricula: The performance of 2000 students in introductory electromagnetism

Matthew A. Kohlmyer,[1,*] Marcos D. Caballero,[2] Richard Catrambone,[3] Ruth W. Chabay,[4] Lin Ding,[5] Mark P. Haugan,[6] M. Jackson Marr,[3] Bruce A. Sherwood,[4] and Michael F. Schatz[2,†]

[1]*School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

[2]*Center for Nonlinear Science and School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

[3]*School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

[4]*Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, USA*

[5]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

[6]*Department of Physics, Purdue University, West Lafayette, Indiana 47907, USA*

(Received 8 October 2008; published 5 October 2009; publisher error corrected 12 October 2009)

The performance of over 2000 students in introductory calculus-based electromagnetism (E&M) courses at four large research universities was measured using the Brief Electricity and Magnetism Assessment (BEMA). Two different curricula were used at these universities: a traditional E&M curriculum and the Matter & Interactions (M&I) curriculum. At each university, postinstruction BEMA test averages were significantly higher for the M&I curriculum than for the traditional curriculum. The differences in post-test averages cannot be explained by differences in variables such as preinstruction BEMA scores, grade point average, or SAT Reasoning Test (SAT) scores. BEMA performance on categories of items organized by subtopic was also compared at one of the universities; M&I averages were significantly higher in each topic. The results suggest that the M&I curriculum is more effective than the traditional curriculum at teaching E&M concepts to students, possibly because the learning progression in M&I reorganizes and augments the traditional sequence of topics, for example, by increasing early emphasis on the vector field concept and by emphasizing the effects of fields on matter at the microscopic level.

DOI: 10.1103/PhysRevSTPER.5.020105 PACS number(s): 01.40.Fk, 01.40.gb

I. INTRODUCTION

Each year more than 100 000 students take calculus-based introductory physics at colleges and universities across the U.S. Such students must obtain a good working knowledge of introductory physics because physics concepts underpin the content of many advanced science and engineering courses required for the students' degree programs. Unfortunately, many students do not acquire an adequate understanding of basic physics from the introductory courses; rates of failure and withdrawal in these courses are often high, and a large body of research has shown that student misconceptions about physics persist even after instruction has been completed.[1] In recent years, there have been significant efforts to reform introductory physics instruction.[2–4]

Reforms of the course content (curricula) of introductory physics have not progressed as rapidly as reforms of content delivery methods (pedagogy). Most students are taught introductory physics in a large lecture format; the shortcomings of passive delivery of content in this venue are well known.[5] A number of pedagogical modifications that improve student learning[2–4] have been introduced and are in widespread use; these modifications range from increasing active engagement of students in large lectures (e.g., peer instruction[4] and the use of personal response system "clickers"[6]) to reconfiguring the instructional environment.[7] By contrast, most students learn introductory physics following a canon of topics that has remained largely unchanged for decades regardless of the textbook edition or authors. As a result the impact of changes in introductory physics curricula on improving student learning is not well understood.

At many universities and colleges, the introductory physics sequence consists of a one-semester course with a focus on Newtonian mechanics followed by a one-semester course in electromagnetism (E&M). There exist a number of standardized multiple-choice tests that can be used to assess objectively and efficiently student learning in large classes of introductory mechanics; some of these instruments have gained widespread acceptance and have been used to gauge the performance of thousands of mechanics students in educational institutions across the U.S.[8] By contrast, fewer such standardized instruments exist for E&M and no single E&M assessment test is widely used. As a result, relatively few measurements of student learning in large lecture introductory E&M have been performed.

In this paper we report measurements of the performance of 2537 students in introductory E&M courses at four large institutions of higher education: Carnegie Mellon University (CMU), Georgia Institute of Technology (GT), North Carolina State University (NCSU), and Purdue University (Purdue). Two different curricula are evaluated: a traditional curriculum, which for our purposes will be defined by a set of similarly organized textbooks in use during the study[9] and the Matter & Interactions (M&I)[10] curriculum. M&I differs from the traditional calculus-based curriculum in its emphasis on fundamental physical principles, microscopic models of matter, coherence in linking different domains of physics, and computer modeling.[11–13] In particular, M&I revises the learning progression of the second semester introductory electromagnetism course by reorganizing and augmenting the traditional sequence of topics, for example, by increasing early emphasis on the vector field concept and by emphasizing the effects of fields on matter at the microscopic level.[14] Student performance is measured using the Brief Electricity and Magnetism Assessment (BEMA) a 30-item multiple-choice test which covers basic topics that are common to

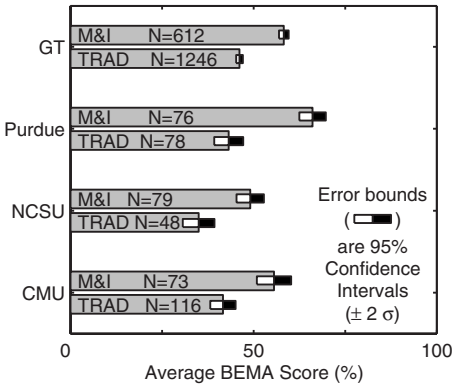

FIG. 1. Average postinstruction BEMA scores at four academic institutions—the average BEMA test scores are shown for students who have completed a one-semester E&M course with either the traditional (TRAD) or M&I curriculum. The number of students tested for each curriculum at each institution is indicated in the figure. The error bounds represent the 95% confidence intervals on the estimate of the average score.

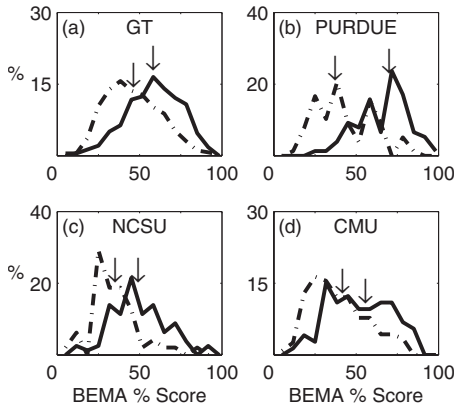

FIG. 2. Postinstruction BEMA score distributions at four academic institutions—the percentage of students with a given BEMA test score is plotted for students who have completed an E&M course with either a traditional (dotted-dashed line) or M&I curriculum (solid line) at (a) GT, (b) Purdue, (c) NCSU, and (d) CMU. The arrows indicate the location of the average score for each distribution. The rightmost arrow in each subfigure corresponds to the M&I course. The total number of students tested in each curriculum at each institution is the same as in Fig. 1. The plots are constructed from binned data with bin widths equal to approximately 6.7% of the maximum possible BEMA score (100%).

both the traditional and M&I electromagnetism curricula including basic electrostatics, circuits, magnetic fields and forces, and induction.[15] In the design of the BEMA, many instructors of introductory and advanced E&M courses were asked to judge draft questions to ensure that questions included on the test did not favor one curriculum over another. Moreover, careful evaluation of the BEMA suggests the test is reliable with adequate discriminatory power for both traditional and M&I curricula.[16]

The paper is organized as follows. In Sec. II we present a summary of BEMA results across the four institutions which provides a snapshot of the performance measurements for students in both the traditional and M&I curricula. In Secs. III–VI we then discuss the detailed results from each individual institution in turn. In Sec. VII we analyze BEMA performance by individual item and topic, discussing possible reasons for performance differences, and we make concluding remarks and outline possible future research directions in Sec. VIII.

II. SUMMARY OF COMMON CROSS-INSTITUTIONAL TRENDS

Comparison of student scores on the BEMA at all four academic institutions suggests that students in the M&I curriculum complete the E&M course with a significantly better grasp of E&M fundamentals than students who complete E&M studies in a traditional curriculum (Figs. 1–3). (A description of the methodology used to define "significance" is given in the Appendix). Broadly speaking, the profiles of students at all institutions were similar; the vast majority of students in both curricula were engineering and/or natural science majors. During the term, all students at a given institution were exposed to an instructional environment with similar boundary conditions on contact hours: large lecture sections that met for 2–4 h/week (depending on the institution) in conjunction with smaller laboratory and/or recitation sections that typically met for 1–3 h/week on average (again depending on the institution). We emphasize that, at a given institution, the contact hours were, for the most part, very similar for both M&I and traditional courses (see Secs. III–VI). Both the average BEMA scores (Fig. 1) and the BEMA score distributions (Fig. 2) were obtained at all institutions by administering the BEMA after students completed their respective E&M courses.

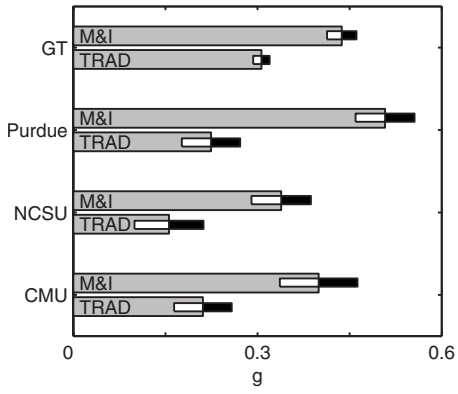

FIG. 3. Gain in student understanding of E&M at four academic institutions—the increase in student understanding resulting from a one-semester traditional (TRAD) or M&I course is measured using the average normalized gain g. The numbers of students tested for each curriculum at each institution: GT M&I: $N=297$; GT Trad.: $N=887$; Purdue M&I: $N=76$; Purdue Trad.: $N=79$; NCSU M&I: $N=79$; NCSU Trad.: $N=48$; CMU M&I: $N=73$; and CMU Trad.: $N=116$. The error bounds represent the 95% confidence intervals on the estimate of the normalized gain. The estimates of g require the average BEMA scores for incoming students I; for the NCSU and CMU results, I was computed differently than for the GT and Purdue results (see Secs. II, V, and VI for details).

TABLE I. Georgia Tech BEMA test results are shown for five Matter & Interactions sections (M1–M5) and 11 traditional sections (T1–T11). Different lecturers are distinguished by a unique letter in column *L*. (Note that lecturer B in M3 was assisted by lecturer A.) The average BEMA score O for N_O students completing the course is shown for all sections. Moreover, in those sections where data are available, the average BEMA score I for N_I students entering the course is indicated. N_m is the number of students in a given section who took the BEMA both at the beginning and at the end of their E&M course. GPA is the incoming cumulative grade point average for students in a given section.

ID	L	N_I	I (%)	N_O	O (%)	N_m	GPA
M1	A	43	24.5 ± 2.3	40	59.8 ± 4.8	40	2.96 ± 0.18
M2	A	n/a	n/a	149	59.7 ± 2.8	n/a	2.99 ± 0.10
M3	B	n/a	n/a	146	57.4 ± 2.6	n/a	n/a
M4	C	138	27.7 ± 1.9	138	59.5 ± 2.7	132	3.14 ± 0.10
M5	D	140	24.7 ± 1.4	139	55.9 ± 2.9	131	3.07 ± 0.09
T1	E	231	22.9 ± 1.2	204	41.2 ± 1.9	180	3.10 ± 0.07
T2	E	219	22.9 ± 1.3	195	40.7 ± 1.9	176	2.99 ± 0.08
T3	F	203	25.7 ± 1.4	136	51.9 ± 3.0	130	3.01 ± 0.09
T4	F	212	25.1 ± 1.4	144	50.8 ± 2.5	133	2.98 ± 0.09
T5	E	n/a	n/a	144	38.3 ± 2.5	n/a	3.09 ± 0.08
T6	G	n/a	n/a	29	45.2 ± 6.5	n/a	2.98 ± 0.12
T7	G	n/a	n/a	36	44.5 ± 4.9	n/a	2.81 ± 0.12
T8	H	87	28.1 ± 2.0	73	54.8 ± 4.7	59	2.97 ± 0.13
T9	J	112	26.5 ± 2.1	84	51.6 ± 3.7	75	2.94 ± 0.11
T10	F	128	25.3 ± 1.6	103	50.3 ± 3.0	88	3.04 ± 0.09
T11	F	127	25.8 ± 1.8	98	49.5 ± 3.3	82	3.03 ± 0.10

A measure of the gain in student understanding as a result of instruction can be obtained by also administering the BEMA to students as they enter the course. Specifically, the average increase in student understanding is measured by the average percentage gain, $G=O-I$, where I is the average BEMA percentage score for students entering an E&M course and O is the average end-of-course BEMA percentage score. It has become customary[8] to report an average normalized gain g, where $g=G/(100-I)$ and $(100-I)$ represents the maximum possible percentage gain that could be obtained by a class of students with an average incoming BEMA score of I. For g reported in Fig. 3, the Georgia Tech and Purdue data are shown only for students who took the BEMA both upon entering and upon leaving their E&M course. For the NCSU and CMU students in this study, I was not measured. In these cases, we estimate g using measurements of I for other similar student populations at each institution (see Secs. V and VI for details on, respectively, the NCSU and CMU estimates). With these qualifications, the data (Fig. 3) show at all four academic institutions that students receiving instruction in the M&I curriculum show significantly greater gains in understanding fundamental topics in E&M than students who received instruction in a traditional curriculum.

As we will discuss later, students who get A's in the course do better on the BEMA than those who get B's, who in turn do better than those who get C's. Comparison of average BEMA scores for a given final course grade in E&M at CMU, NCSU, and GT suggests that, roughly speaking, M&I students perform one-letter grade higher than students in the traditional-content course. For example, on average an M&I student with a course grade of B does as well on the BEMA as the traditional-content student with a course grade of A.

In addition to the common features described here, the E&M instructional and assessment efforts contained a number of details unique to each academic institution. We discuss these details below (Secs. III–VI).

III. GEORGIA TECH BEMA RESULTS

The typical introductory E&M course at Georgia Tech is taught with three 1 h lectures/week in large lecture sections (150–250 students per section) and 3 h/week in small-group (20 student) laboratories and/or recitations. In the traditional curriculum, each student attends a 2 h laboratory and, in a separate room, a 1 h recitation each week; in the M&I curriculum, each student meets once per week in a single room for a single 3 h session involving both laboratory activities (for approximately 2 h on average) and separate recitation activities (for approximately 1 h on average). The student population of the E&M course (both traditional and M&I) consists of 83% engineering majors and 17% science (including computer science) majors.

Table I summarizes the Georgia Tech BEMA test results for individual sections. In all traditional and M&I sections, N_O students in each section took the BEMA during the last week of class at the completion of the course, typically during the last lecture or laboratory session. Moreover, in the majority of both traditional sections (T1-T4 and T8-T11) and

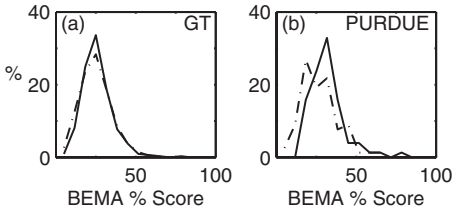

FIG. 4. Pretest BEMA score distributions for Georgia Tech and Purdue—the distributions of BEMA test scores for students before completing an E&M course with either a traditional (dotted-dashed line) or M&I curriculum (solid line) are shown for data from (a) GT ($N=1319$ for traditional students; $N=321$ for M&I students) and (b) Purdue ($N=78$ for traditional students; $N=76$ for M&I students). The plots are constructed from binned data with bin widths equal to approximately 6.7% of the maximum possible BEMA score (100%).

M&I sections (M1, M4, and M5), N_I students in each section took the BEMA at the beginning of the course during the first week of class, typically during the first lecture or laboratory section. N_I for a given section is approximately equal to the number of students enrolled in that section, while N_O is usually smaller than N_I, sometimes substantially so (e.g., T3 and T4), due to the logistics of administering the test. Thus, in each section, only those N_m students who took the BEMA both on entering and on completion of the course are considered for the purposes of computing both the un-normalized gain G and the normalized gain g. The BEMA was administered using the same time limit (45 min) for both traditional and M&I students. M&I students were given no incentives for taking the BEMA; they were asked to take the exam seriously and told that the score on the BEMA would not affect their grade in the course. Traditional students taking the BEMA were given bonus credit worth up to a maximum of 0.5% of their final course score, depending in part on their performance on the BEMA. A performance incentive for only traditional students would not be expected to contribute to poorer performance of traditional students relative to M&I students and, therefore, cannot explain the Georgia Tech differences in performance summarized in Figs. 1 and 2.

Figure 4(a) demonstrates that there was no significant difference between traditional and M&I students in the distribution of pretest scores on the BEMA. The average pretest score for all sections ranged from about 22% to 28%; a section-by-section comparison suggests there is no significant difference in pretest scores on the BEMA between individual sections (see Table I and Appendix). As an additional check on student populations in the two curricula, we examined the students' grade point averages (GPA) at the start of the E&M courses; no significant difference in incoming GPA was found.[17] Thus, the student population entering both courses is essentially the same. Additionally, because the BEMA pretest averages and the distribution of BEMA pretest scores are essentially the same for the GT students in both curricula, we focus our remaining discussion on the post-test scores.

Figure 2(a) indicates that the distribution of the BEMA post-test scores for the M&I group is significantly different than the distribution for the traditional group. Moreover, the

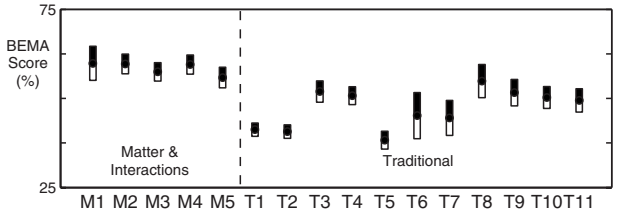

FIG. 5. Average BEMA scores by section at Georgia Tech—the average end-of-semester BEMA scores for 11 traditional (T#) and 5 M&I (M#) sections at GT are shown. The error bounds indicate 95% confidence intervals on the estimates of the average for each section. The number of students tested in a particular section is given by N_m in Table I.

BEMA post-test averages for each section (Fig. 5) suggest that the M&I sections consistently outperform the traditional sections. The M&I BEMA averages across four different instructors are relatively consistent, while the BEMA averages of the traditional sections across five different instructors vary greatly. The use of personal response system clicker questions may account for some of this difference. The lowest scoring sections (T1, T2, and T5 in Fig. 5) did not use clicker questions; by contrast, approximately two to six clicker questions were asked per lecture in all M&I sections and all other traditional sections. Nevertheless, even when the comparison between sections is restricted to the traditional sections with the highest average BEMA scores (sections T3, T4, T8, and T9, which were taught by three different instructors who have a reputation of excellent teaching), the M&I sections demonstrated significantly better performance (Appendix).

The data in Fig. 6 suggest a correlation between BEMA scores and final course grade at GT, with M&I students outperforming traditional students with the same final letter grade. Our finding that BEMA scores correlate strongly with final letter grade is not obvious. It seemed possible that the course grade was determined to a significant extent by the students' ability to work difficult multistep problems on exams, whereas the BEMA primarily measures basic concepts

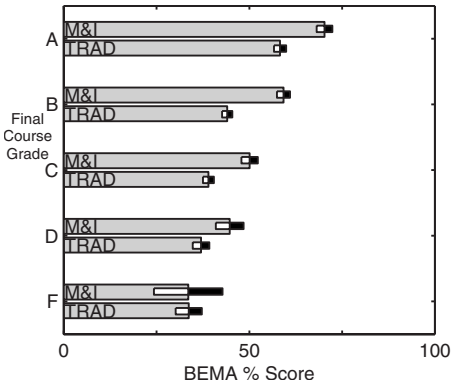

FIG. 6. The average post-test BEMA score of all students receiving a particular final course grade in introductory E&M at GT is shown. The error bounds indicate 95% confidence intervals on the estimates of the average for each grade. The numbers of students for whom grades were obtained are $N=1233$ for traditional students and $N=611$ for M&I students.

which, it was hoped, all students would have mastered. However, we find M&I students exhibit a one-letter-grade performance improvement as compared with traditional students; specifically, the average BEMA scores are statistically equivalent between traditional A students and M&I B students, traditional B students and M&I C students, and traditional C students and M&I D students. This difference in performance cannot be attributed to differences in the distribution of final grades; the percentage of students receiving a given final grade in the M&I sections (27.7% A's, 37.8% B's, 25.2% C's, 7.2% D's, and 2.1% F's) is similar to that in the traditional sections (29.8% A's, 34.4% B's, 24.3% C's, 8.8% D's, and 2.7% F's).

IV. PURDUE BEMA RESULTS

The curriculum comparison at Purdue focuses on an introductory E&M course taught to electrical and computer engineering majors. The contact time was allocated somewhat differently for students in each curriculum; however, the total course contact time was similar for both traditional and M&I students. Each week, traditional students met for three 50 min large lectures (approximately 100 students per section) and two 50 min small-group recitations (25–30 students); these students did not attend a laboratory. M&I students met for two 50 min lectures/week in large lecture sections (approximately 100 students per section) and 2 h/week in small-group (25–30 students) laboratories. In addition, M&I students attended a small-group (25–30 students) recitation once a week for 50 min. In all traditional and M&I sections, students in each section took the BEMA during the last week of class at the completion of the course, typically during the last lecture or laboratory session. Moreover, students in each section took the BEMA at the beginning of the course during the first week of class, typically during the first lecture or laboratory section.

Figure 2(b) indicates M&I students significantly outperformed traditional students at Purdue. Students in both courses took the BEMA during a portion of a laboratory period with a 45 min time limit for completion. Both traditional and M&I students took the assessment (both pre and post) in the same week. The "initial state" of the two groups upon entering their respective E&M course was measured by comparison of the grade point averages between the two classes; no significant difference was found.[18] Additionally, comparison of the distributions of the BEMA score upon entrance to the course shows only a small difference between the two groups [Fig. 4(b)] that cannot account for the large post-test difference shown in Fig. 2(b).

V. NORTH CAROLINA STATE BEMA RESULTS

The introductory E&M course at North Carolina State is typically taught with three 1 h lectures/week in large lecture sections (about 80 students per section). (Note, however, that one M&I section was taught in the SCALE-UP studio format.[7]) In the traditional curriculum, each student attended a 2 h laboratory every two weeks; in the M&I curriculum, each student attends a 2 h laboratory every week. Approximately three-fourths of the student population of the E&M course (both traditional and M&I) are engineering majors.

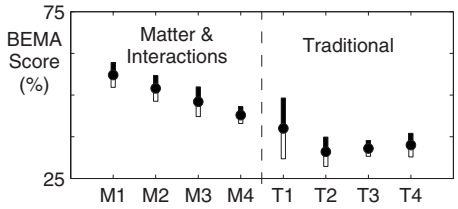

FIG. 7. Average BEMA score by section at NCSU—the average end-of-semester BEMA scores for four traditional (T#) and four M&I (M#) sections at NCSU are shown. The error bars indicate 95% confidence intervals on the estimates of the average for each section. The numbers of students tested: $N=7$ for T1, $N=10$ for T2, $N=16$ for T3, $N=15$ for T4, $N=16$ for M1, $N=22$ for M2, $N=10$ for M3, and $N=31$ for M4. Note that section M4 was taught in the SCALE-UP studio format.

One hundred twenty-seven volunteers were recruited from eight different sections (a total of 700 students) by means of an in-class presentation made by a physics education research graduate student. Students were paid $15 for their participation in this out-of-class study. Prior to participation, students were told that they did not need to study for the test. Just before the end of the semester, several testing times were scheduled to accommodate student schedules. The test was given in a classroom containing one computer per student, with a proctor present; each student took the test using an online homework system. Each student took the test independently with a 60 min time limit.

The difference in BEMA averages (shown in Fig. 1) between the M&I group and the traditional group is large and statistically significant as determined by the method outlined in the Appendix. Because students were recruited from eight different sections, it is of interest to observe how students from each section performed on the BEMA. Figure 7 shows the average scores of the individual sections for both M&I and traditional groups. Results of statistical tests (namely, the Kruskal-Wallis test[19]) show that there was no significant difference in BEMA scores among the M&I sections; similarly, no significant difference across the four traditional sections was detected. These results suggest that within each group students' BEMA scores were statistically uniform and that the better performance of the M&I students was not due to a few outlier sections that could have biased the results.

One possible explanation of the results may be a recruitment bias; that is, higher-performing M&I students and lower-achieving traditional students may have been recruited for the study. To rule this out, participants' GPA, SAT Reasoning Test (SAT) scores, as well as math and physics course grades (prior to taking the E&M course) were examined. The two math courses from which students' grades were collected were the first and second semesters of calculus courses; the physics course for which students' grades were collected was the calculus-based mechanics course. Using the method described in the Appendix, we found that there was no significant difference between the M&I group and traditional group in any of these grades. Additionally, no significant difference was found in the SAT scores (verbal and math scores). These

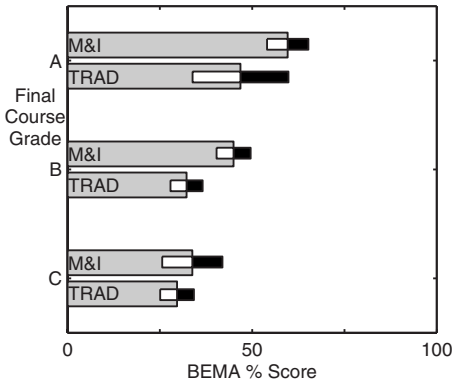

FIG. 8. The average post-test BEMA score of all students receiving a particular letter grade in introductory E&M at NCSU is shown. The error bounds indicate 95% confidence intervals on the estimates of the average for each grade. The numbers of students for which grades were obtained are $N=48$ for traditional students and $N=79$ for M&I students.

results suggest that the recruitment was not biased and that student participants from both the M&I sections and traditional sections had similar academic backgrounds.[20]

In the NCSU study, students were not given the BEMA prior to the start of their E&M course. However, a number of students from the same population, who were concurrently enrolled in introductory mechanics, did take the BEMA using via a web-based delivery system. The average BEMA score of the mechanics students was 23%.[21] We use this value as an estimate for I to compute the normalized gains shown in Fig. 3 which shows superior gain by M&I students.

The data in Fig. 8 suggest a correlation between BEMA scores and final course grade at NCSU, with M&I students outperforming traditional students with the same final letter grade. Moreover, we find that M&I students exhibit a one-letter-grade performance improvement; specifically, the average BEMA scores are statistically equivalent between traditional A students and M&I B students. Such a performance difference might arise if fewer high final grades were awarded in M&I than in the traditional course; under these circumstances, the A students in M&I would be more select and, perhaps, better than A students in the traditional course. In fact, however, a somewhat larger percentage of higher final grades was earned in the M&I sections (40.5% A's, 43.0% B's, 12.7% C's, 2.5% D's, and 1.3% F's) than in the traditional sections (25.0% A's, 54.2% B's, 18.7% C's, 2.1% D's, and 0.0% F's). Thus, the difference in performance on the BEMA cannot be attributed to differences in the distribution of final grades.

VI. CARNEGIE MELLON RETENTION STUDY

The introductory E&M course at Carnegie Mellon consisted of a large (~150 students) lecture that met 3 h/week and a recitation section that met 2 h/week; there was no laboratory component to this course. For historical reasons, the course was separated into two versions: one for engineering majors that used the traditional curriculum and one for natural and computer science majors that used the M&I curriculum. The pedagogical aspects of both the traditional and M&I courses were quite similar.

To probe the retention of E&M concepts as a function of time, two groups of students were recruited from each curriculum: (1) recent students of introductory E&M, i.e., students who had taken the introductory E&M final exam 11 weeks prior to BEMA testing and (2) "old" students of introductory E&M, who had completed introductory E&M anywhere from 26 to 115 weeks prior to BEMA testing. A total of 189 students volunteered for the study out of a pool of 1200 CMU students who had completed introductory E&M at CMU and who were sent a recruitment email by a staff person outside the physics department. With a promise of a $10 honorarium, the electronic mail asked for volunteers to take a retention test on an unspecified subject and stated that the test's purpose was to contribute to improvement in introductory courses. The student volunteers took the BEMA during the evening in a separate proctored classroom. Just before taking the test, students were again told that they could help improve instruction at CMU by participating and doing their best; a poll of the students indicated, with one exception, that the volunteers arrived at the examination room without knowledge of the test's subject matter. No pretest was given to the students; however, an estimate of I, the average BEMA score prior to entering the E&M course, was obtained by a separate study. To obtain this estimate, a different group of volunteers drawn from the appropriate pool of potential students for each curriculum, i.e., engineering students who had not yet taken the traditional E&M course and science students who had not yet taken the M&I E&M course. These volunteers were given the BEMA; we estimate $I=28\%$ ($N=14$) for the traditional courses and $I=23\%$ ($N=10$) for M&I.

Disregarding the length of time since completing the E&M course, it was found that the average BEMA score $O=41.6\%$ for students in the traditional curriculum is significantly lower than the $O=55.6\%$ for students in the M&I curriculum. The participants from each course were not significantly different in background as measured by the average SAT verbal or math score.

Figure 9 shows that E&M knowledge as measured by the BEMA showed a significant loss over the retention period for both M&I and traditional students. While the M&I groups showed greater absolute retention at all grade levels than the traditional groups, the BEMA performances of students who most recently completed the E&M course were also greater in the M&I group. The rate of loss in the two groups appeared to be the same, a result typically found in the experimental analysis of retention when comparing different initial "degrees of learning."[22,23] Thus, as measured by BEMA performance we could not determine unequivocally that M&I improved retention of E&M knowledge over the traditional course beyond effects due to initial differences in performance on the BEMA. It is worth noting here that recent work has shown that better retention occurs for students exposed to improved pedagogical techniques.[24]

The data in Fig. 10 suggest a correlation between BEMA scores and final course grade at CMU, with M&I students outperforming traditional students with the same final letter grade. Moreover, we find that M&I students exhibit a one-

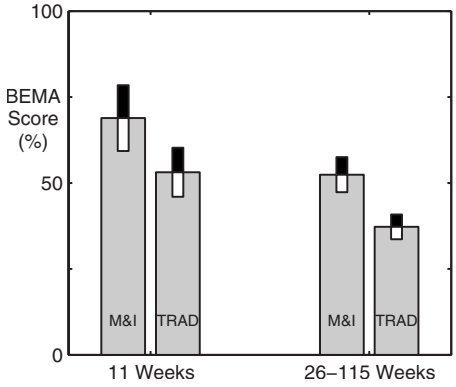

FIG. 9. Retention of E&M knowledge—average scores on the BEMA vs time since completion of a course in introductory E&M are shown for students at CMU from either a traditional curriculum (TRAD) or the M&I curriculum. The error bounds indicate 95% confidence intervals on the estimates of the average for each section. The numbers of students tested were 116 for traditional and 73 for M&I.

letter-grade performance improvement; specifically, the average BEMA scores are statistically equivalent between traditional A students and M&I B students. Such a performance difference might arise if fewer high final grades were awarded in M&I than in the traditional course; under these circumstances, the A students in M&I would be more select and, perhaps, better than A students in the traditional course. In fact, however, a somewhat larger percentage of higher final grades was earned in the M&I sections (34.3% A's, 39.7% B's, 21.9% C's, 4.1% D's, and 0% F's) than in the traditional sections (25.0% A's, 37.9% B's, 31.9% C's, 5.2% D's, and 0.0% F's). Thus, the difference in performance on the BEMA cannot be attributed to differences in the distribution of final grades.

VII. ITEM ANALYSIS OF THE BEMA

We have seen superior performance on the BEMA from M&I introductory E&M classes as compared to traditional

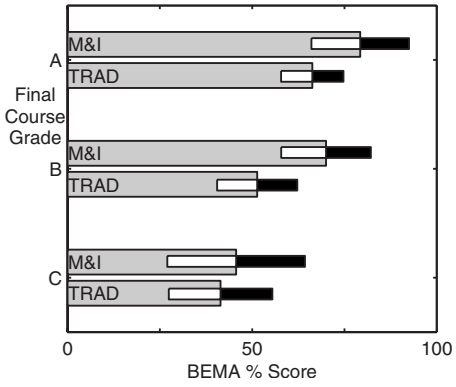

FIG. 10. The average BEMA score for students receiving a particular final grade in introductory E&M at CMU is shown. In all cases, the BEMA test was administered 11 weeks after the completion of the course. The error bounds indicate 95% confidence intervals on the estimates of the average for each grade. The numbers of students for which grades were obtained are $N=14$ for traditional students and $N=32$ for M&I students.

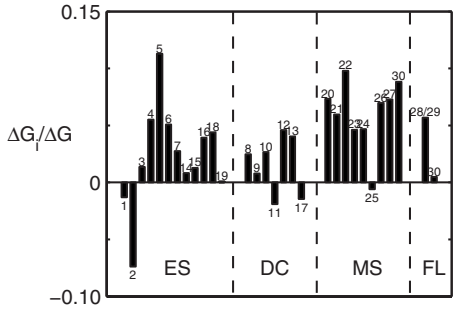

FIG. 11. Fractional difference in performance for E&M subtopics—the fractional difference in performance $\Delta G_i/\Delta G$ between M&I and traditional students at GT is shown for each question on the BEMA. Positive (negative) $\Delta G_i/\Delta G$ indicates superior performance by M&I (traditional) students. The numerical labels indicate the corresponding question number in order of appearance on the BEMA. The $\Delta G_i/\Delta G$ are grouped together into one of the four topics: electrostatics (ES), dc circuits (DC), magnetostatics (MS), or Faraday's law and induction (FL).

E&M classes across multiple institutions. One question that arises is whether this result can be explained by M&I students performing better in any one topic or set of topics in the E&M curriculum. Because the content of the BEMA spans a broad range of topics, we can examine this question by dividing the individual BEMA items into different categories and comparing M&I and traditional course performances in the individual categories. There are some subjective decisions to be made when categorizing the items based on content and concepts, including the number of categories, the particular concepts they encompass, and which items belong to which categories. Furthermore, certain items may involve more than one concept and could potentially fall into more than one category. We decided, for simplicity, to group the BEMA items into just four categories covering different broad topics, namely, electrostatics, dc circuits, magnetostatics, and Faraday's law of induction. Each item was placed into one and only one category; refer to Fig. 11 for the items that comprise each category.[25] Note that this is an *a priori* categorization based on physics experts' judgment of the concepts covered by the items; it is not the result of internal correlations or factor analysis based on student data. Using these categories, we compared M&I and traditional performance in each category. We chose to analyze the data from GT only because we had the largest amount of data for traditional and M&I courses across a range of different lecture sections from this institution.

We define the difference in performance between the two curricula as $\Delta G=G_M-G_T$, where G_M and G_T are the (unnormalized) gains for the M&I and traditional curricula, respectively. In the same way, we can determine ΔG_i as the difference in performance of the ith BEMA question; ΔG_i is equal to the percentage of M&I students that answered the ith question correctly minus the percentage of traditional students that answered the same question correctly. Using these quantities, we define $\Delta G_i/\Delta G$ as the fractional difference in performance for the ith question. $\Delta G_i/\Delta G$ can be thought of as the fractional contribution of the ith question to ΔG since $\Sigma_i \Delta G_i/\Delta G=1$. For equal weighting in the BEMA score (the

scoring method that we used), a given question will make an "average" contribution to ΔG when the magnitude of $\Delta G_i/\Delta G$ is approximately equal to the inverse of the number of test questions (0.033 for the 30-question BEMA). Thus, when the magnitude of $\Delta G_i/\Delta G$ is significantly greater than 0.033, the corresponding question yields a greater than average contribution to ΔG. In addition, the sign of $\Delta G_i/\Delta G$ is noteworthy; a positive (negative) $\Delta G_i/\Delta G$ corresponds to an item where on average the M&I students scored higher (lower) than traditional students. (This presumes $\Delta G > 0$, which is the case for our data.)

The plot of $\Delta G_i/\Delta G$ for all questions on the BEMA provides a kind of "fingerprint" for comparing in detail the performance of M&I and traditional students (Fig. 11). We see that the M&I course has positive $\Delta G_i/\Delta G$ for almost all questions on the BEMA, and more than half of the questions (16) have values of $\Delta G_i/\Delta G$ greater than 0.033.[26] The grouping of the BEMA questions by category permits one to visualize which topics contribute most strongly to the difference in performance. For example, the difference in performance in magnetostatics is striking, where nearly every question in this category has $\Delta G_i/\Delta G > 0.033$; in fact M&I student performance on magnetostatics alone accounts for more than half (55%) of the difference in performance ΔG relative to traditional students. The positive $\Delta G_i/\Delta G$ for dc circuits is worthy of note even though these questions account for only 12% of ΔG. Qualitatively speaking, the M&I course seeks to connect the behavior of circuits to the behavior of both transient and steady-state fields; this focus is decidedly nontraditional. By contrast, the dc-circuit questions on the BEMA are quite traditional, so it is tempting to think that the traditional course might provide better training for responding to such questions. However, Fig. 11 demonstrates that in fact M&I students outperform traditional students on traditional dc-circuit questions. Performance in electrostatics also generally favors the M&I course (28% contribution to ΔG); however, we see that the performance on question 2 significantly favors the traditional course. The topic of question 2 is the computation of electric forces using Coulomb's law. It is possible that the difference is due to greater time spent in the traditional class on electric forces between point charges at the beginning of the course. The M&I curriculum also discusses forces on charges but moves into a full discussion of electric fields due to point charges more quickly than the traditional course, thereby devoting less time to discussing forces exclusively. By contrast, we also see the largest single percentage difference in favor of M&I in question 5, which deals with the direction of electric field vectors due to a permanent electric dipole. The electric dipole plays an important role in the M&I curriculum due to the curriculum's emphasis on the effects of electric fields on solid matter and polarization, topics which are often skipped or deemphasized in the traditional course; this particular result is therefore not particularly surprising. As a final note, the large values of $\Delta G_i/\Delta G$ between M&I and traditional courses in both magnetostatics and Faraday's law are interesting because these topics are regarded as the most difficult for students due to their high level of abstraction and geometric complexity. It is therefore striking that the M&I curriculum seems to be making the largest impact on the hardest topics at least at Georgia Tech.

As an independent check on the significance of our item analysis, we used the method of contingency tables as described in the Appendix to compare the M&I and traditional students' average scores in each individual category. Here, a student's score in a category is computed as the sum of correct items in that category, where the number of items in the four categories ranges from 2 to 12. The discrete nature of the data, as well as the non-normality and unequal variances of the distributions, makes contingency tables the appropriate choice for this type of analysis. On the pretest, we found no significant association between course treatment (M&I versus traditional) and overall BEMA score on any category. In contrast, the results of the contingency table analysis (see the Appendix) for the post-test scores show significant association of BEMA score with treatment in each category. We interpret this as showing better performance across topics for students in the M&I course.

VIII. DISCUSSION

We have presented evidence that introductory calculus-based E&M courses that use the Matter & Interactions curriculum can lead to significantly higher student postinstruction averages on the Brief Electricity and Magnetism Assessment than courses using the traditional curriculum. The strength of this evidence is bolstered by the number of different institutions where this effect is measured and by the large number of students involved in the measurements. We interpret these results as showing that M&I is more effective than the traditional curriculum at providing students with an understanding of the basic concepts and phenomena of electromagnetism. This interpretation is based on accepting that the BEMA is a fair and accurate measurement of such an understanding. We believe that this is a reasonable proposition with which most E&M instructors would agree, given that the BEMA's items cover a broad range of topics common to most introductory E&M courses. However, the BEMA was designed to measure just this minimal subset of common topics. There may be other topics in which traditional students would outperform M&I because they are not taught or de-emphasized in the M&I course and vice versa.

The BEMA is not the only instrument to assess student understanding of E&M concepts. The Conceptual Survey of Electricity and Magnetism (CSEM) was also designed for such a purpose.[27] With the exception of electric circuits, omitted by the CSEM, both instruments cover similar topics; in fact, several items are common to both tests. However, the CSEM contains questions involving field lines, a topic which is not covered in the M&I curriculum (a justification for this omission is discussed in Ref. [14]). Recent work has shown the CSEM and BEMA to be equivalent measures for changes in pedagogy.[28] Nevertheless, it would be interesting to use the CSEM in comparative assessments of traditional and M&I courses to see if it gives results similar to the BEMA; several of us are planning to do this in future semesters.

A major research question raised by these results is how and why the M&I curriculum is leading to higher performance on the BEMA. The postinstruction BEMA results measure only the total effect of the content and pedagogy of

the entire course; there is no way to tease out from these measurements the effects of any individual elements of a course. While it is true that interactive instruction methods (clickers) were used in almost every M&I class measured, they were also used in many of the traditional classes. Recall that M&I sections at Georgia Tech still outperformed traditional sections with the two instructors noted for excellent pedagogical techniques. Overall performance differences are not likely to be explained by differences in overall time on task; the weekly classroom contact time was equivalent for both M&I and traditional students at two of the four institutions (Georgia Tech and Purdue). Time on task for specific E&M topics may partially explain performance differences like those shown in Fig. 11. Comparing the percentage of total lecture hours devoted to each topic at GT, we find the M&I course spends significantly more lecture time than the traditional course on magnetostatics (24% vs 12%); this is consistent with the superior performance of M&I students on this topic. However, we find superior performance of M&I students on electrostatics, for which both courses spend nearly equal lecture time (36% vs 38%). In addition, we also find superior M&I performance of topics where the M&I course spends significantly less lecture time than the traditional course, namely, dc circuits (15% vs 25%) and Faraday's law and induction (6% vs 11%). We conclude that topical time on task alone is insufficient to account for performance differences on the BEMA.

It is possible that the revised learning progression offered by the M&I E&M curriculum is responsible for the higher performance on the BEMA by M&I students. For example, more time is spent exclusively on charges and fields early in the course, laying conceptual groundwork for the mathematically more challenging topics of flux and Gauss's law which are dealt with later than is traditional. Also, magnetic fields are introduced earlier than is traditional, giving students more time to master this difficult topic. Finally, M&I emphasizes the effects of fields on matter at the microscopic level. In some of the traditional courses discussed in this paper, dipoles and polarization are not discussed.

ACKNOWLEDGMENTS

We would like to thank Deborah Bennett, Lynn Bryan, Dan Able, and Melissa Yale for their assistance in collecting the Purdue BEMA data. We acknowledge Andrew Scherbakov, Martin Jarrio, and Eric Murray for their efforts in collecting BEMA data from traditional course students at Georgia Tech. We also acknowledge Robert Hume in the Office of Minority Education Development at Georgia Tech for his work in compiling student demographic data. Finally, we thank Keith Bujak for his assistance in editing this paper. This work was supported generously by Division of Undergraduate Education, National Science Foundation (Grant No. DUE0618504, No. DUE0618519, and No. DUE0618647).

APPENDIX: HYPOTHESIS TESTING AND CONFIDENCE INTERVALS

In this paper, data comparing the two curricula were subjected to rigorous testing for statistical significance. Details on the statistical testing can be accessed online via a link to supplemental material located on the abstract page for this paper.

*Present address: Department of Physics, North Carolina State University; makohlmy@unity.ncsu.edu

†Corresponding author; michael.schatz@physics.gatech.edu

[1] I. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).

[2] L. McDermott, P. Shaffer, and The Physics Education Group at the University of Washington, *Tutorials in Introductory Physics* (Pearson Education, Upper Saddle River, NJ, 2002).

[3] C. Hieggelke, D. Maloney, S. Kanim, and T. O'Kuma, *E&M TIPERS: Electricity and Magnetism Tasks (Inspired by Physics Education Research)* (Pearson Education, Upper Saddle River, NJ, 2006).

[4] E. Mazur, *Peer Instruction: A User's Manual* (Addison-Wesley, Reading, MA, 1997).

[5] A. C. to the NSF Directorate for Education and H. Resources, 1996.

[6] C. Wieman and K. Perkins, Transforming physics education, Phys. Today **58**(11), 36 (2005).

[7] M. Oliver-Hoyo and R. Beichner, in *Teaching and Learning through Inquiry: A Guidebook for Institutions and Instructors*, edited by V. S. Lee (Stylus, Sterling, VA, 2004).

[8] 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 (1998).

[9] Textbooks used in the traditional E&M courses at the time of evaluation for each institution are the following: R. Knight, *Physics for Scientists and Engineers with Modern Physics: A Strategic Approach* (Addison-Wesley, Reading, MA, 2004); P. Tipler and G. Mosca, *Physics for Scientists and Engineers*, Vol. 1, (W. H. Freeman and Company, New York, 2004) (Purdue); D. Giancoli, *Physics for Scientists and Engineers*, (Prentice Hall, Upper Saddle River, New Jersey, 2000) (NCSU); H. Young and R. Freedman, *University Physics*, (Addison-Wesley, Reading, MA, 1996) (CMU).

[10] Two of the authors of this paper, R. Chabay and B. Sherwood, are the authors of the *Matter & Interactions* textbook: R. Chabay and B. Sherwood, *Matter & Interactions II: Electric & Magnetic Interactions*, (Wiley, New York, 2007), 2nd ed.

[11] R. Chabay and B. Sherwood, Bringing atoms into first-year physics, Am. J. Phys. **67**, 1045 (1999).

[12] R. Chabay and B. Sherwood, Modern Mechanics, Am. J. Phys. **72**, 439 (2004).

[13] R. Chabay and B. Sherwood, Computational physics in the introductory calculus-based course, Am. J. Phys. **76**, 307 (2008).

[14] R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, Am. J. Phys. **74**, 329 (2006).

[15] For a copy of the BEMA, contact the corresponding author.

[16] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).

[17] The mean incoming GPA at Georgia Tech for M&I students was 3.10 on a 4.0 scale, while the mean incoming GPA for traditional students was 3.11.

[18] At Purdue, the mean incoming GPA for M&I students was 3.25 and for traditional students the mean incoming GPA was 3.14.

[19] P. Sprent, *Applied Nonparametric Statistical Methods*, 2nd ed. (Chapman and Hall, London, 1993).

[20] At NCSU, the mean incoming GPA for M&I students was 3.29 on a 4.0 scale, while the mean incoming GPA for traditional students was 3.23. The mean calculus I GPAs were for 3.50 and 3.53 for M&I and traditional students, respectively, and the mean SAT scores were 1246.5 and 1248.3 for M&I and traditional students, respectively.

[21] In fact, incoming students tend to earn similar BEMA scores across institutions regardless of curriculum: Georgia Tech (25.1%), Carnegie Mellon (25.9%), and Colorado (27%).

[22] N. Slamecka and B. McElree, Normal Forgettinig of Verbal Lists as a Function of Their Degree of Learning, J. Exp. Psychol. Learn. Mem. Cogn. **9**, 384 (1983).

[23] J. Wixted and E. Ebbesen, On the form of forgetting, Psychol. Sci. **2**, 409 (1991).

[24] S. J. Pollock, PERC Proceedings, *A Longitudinal study of the Impact of Curriculum on Conceptual Understanding in E&M*, AIP Press, NY, 2008 .

[25] Item 29 of the BEMA is not scored separately—items 28 and 29 together count as one question and must both be answered correctly to receive credit.

[26] These questions are 4–6, 12, 13, 16–18, 20–24, 26, 27, and 28/29.

[27] D. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[28] S. J. Pollock, PERC Proceedings, *Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA*, AIP Press, NY, 2008.