

## Peaks and decays of student knowledge in an introductory E&M course

Eleanor C. Sayre and Andrew F. Heckler

*Department of Physics, The Ohio State University, Columbus, Ohio 43201, USA*

(Received 22 September 2008; published 12 February 2009)

A common format for assessment of learning is pretesting and post-testing. In this study, we collect student test data several times per week throughout a course, allowing for the measurement of the changes in student knowledge with a time resolution on the order of a few days. To avoid the possibility of test-retest effects, separate and quasirandom subpopulations of students are tested on a variety of tasks. We report on data taken in a calculus-based introductory E&M class populated primarily by engineering majors. Unsurprisingly for a traditional introductory course, there is little change in many conceptual questions. However, the data suggest that some student performance peaks and decays rapidly during a quarter, a pattern consistent with memory research yet unmeasurable by pretesting and post-testing. In addition, it appears that some course topics can interfere with prior knowledge, decreasing performance on questions related to earlier topics in the course.

DOI: [10.1103/PhysRevSTPER.5.013101](https://doi.org/10.1103/PhysRevSTPER.5.013101)

PACS number(s): 01.40.Fk, 01.30.lb, 01.50.Kw, 01.40.Ha

### I. INTRODUCTION

In physics education research, a common method to assess instruction is pretesting and post-testing. In large introductory classes, these assessments often take the form of multiple-choice concept inventories or surveys and are administered at the beginning and end of the term. Some popular tests<sup>1-5</sup> have been administered to many students at multiple institutions.<sup>6</sup> Any change in performance on the instrument is generally attributed to student learning related to the class, and conclusions about the value of teaching methods are commonly based (at least in part) on changes in these scores.

Testing at the beginning and end of a course is an important tool for assessing instruction, but it does not address the extent to which there may be significant changes in student performance on time scales shorter or longer than the duration of the course. In fact, some studies have shown that performance on these tests can be sensitive to when the tests are administered. For example, Ding *et al.*<sup>7</sup> demonstrated that pretest scores can vary significantly when the test administration varies only on the order of a few days. On longer time scales, others have demonstrated<sup>8</sup> that student performance can decrease months after a course has ended, though this can depend on the kind of instruction given. The field of cognitive psychology offers some reasons to suspect that student performance during a course is much more complicated than a simple before-and-after picture. In particular, three well-known effects can contribute to changes in student performance on a given instrument: learning, forgetting, and interference. Learning curves, such as those predicted by the influential Rescorla-Wagner model,<sup>9</sup> show that for repeated training, scores will increase quickly at first, then level off. This simple error-reduction model assumes that the amount of learning depends on the difference between perfect performance and actual performance and matches empirical data in a wide array of simple learning tasks in a variety of species, including humans. Forgetting curves, such as those studied by Hermann Ebbinghaus,<sup>10</sup> reveal that memory performance exponentially decays after training ceases, eventually reaching a new minimum value, and have been demonstrated in a

wide variety of tasks and time scales (from seconds to decades). A third phenomenon, interference, occurs when two pieces of related information (or tasks) are learned. Performance on one can significantly decrease when the second is learned either before (proactive) or after (retroactive) the first one is learned, and the amount of interference increases with the degree of similarity between the two pieces of information.<sup>11,12</sup>

In this Brief Report, we report on initial data that demonstrate some of the dynamical features of student performance throughout a physics course. This is part of a larger study on the measurement of student performance on a series of focused questions. This fundamental research can provide higher time resolution data on student learning than pretesting and post-testing. Consequently, it can provide better feedback on the effects of instruction and insight into some fundamental cognitive mechanisms at play during a physics course.

### II. SIMPLE MODEL

To gain insight into how student performance might vary in time during a course, and perhaps eventually to mathematically model the data, we introduce the following toy model which is based on the fundamental cognitive mechanisms of learning, memory decay, and interference.

Figure 1 presents an example of the model. Before an instructional intervention, the students begin at an initial performance level (i.e., pretest score). Once instruction on the topic of interest occurs, the scores begin to rapidly rise.<sup>13</sup> The exponentially asymptotic nature of the growth in performance during instruction is characteristic of error-driven learning models, such as the Rescorla-Wagner model.<sup>9</sup> During this instructional intervention period, the main parameters are the intervention duration, the short term gain in performance, and the inherent rise time of the performance, which characterizes how quickly students learn to answer a particular question correctly. After the instruction of the particular topic ends, the performance decreases exponentially, characteristic of memory decay, relaxing toward a final performance level (i.e., post-test score).

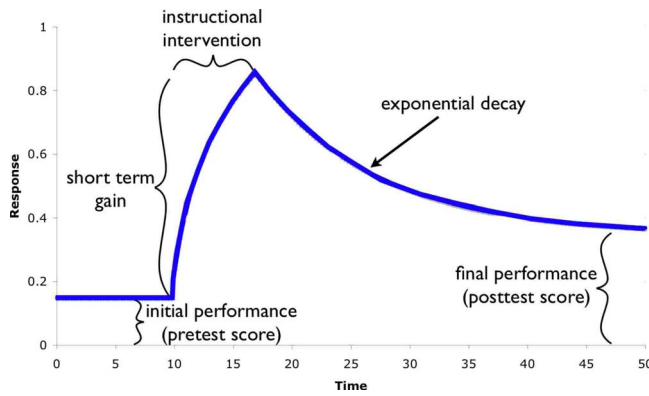


FIG. 1. (Color online) A simulated response to teaching. The curve has four labeled parameters: initial height, short term gain during an instructional intervention, exponential decay constant, and final height.

This model is fully determined with seven numerical parameters; however there are only four values which are of interest here: prescore, postscore, short term gain, and performance decay time. Predata and postdata are already considered in research; we are only introducing two important new parameters—short term gain and decay time—which describe the dynamic character of student performance. These parameters describe the extent to which students may learn to answer a question correctly but subsequently quickly forget.

Since interference depends on the extent to which a second topic is learned, we can also model interference as the negative of the learning-decay curve. Thus interference can occur quickly as the second topic is quickly learned and then disappear with some exponential decay time as the interference resolves or the second topic is forgotten.

### III. EXPERIMENTAL DESIGN

The primary goal for this study is to determine how student performance on a set of specific conceptual questions changes throughout the duration of a course. To avoid any possible test-retest effects such as task familiarity, each student answers any given question only once during the course. Every week, we quasirandomly select a different group of students to answer the same set of questions. We label the selection as quasirandom because we solicit students from two or three different laboratory sections every week via in-laboratory sign-up sheets and electronic mail. Each student signs up for a particular day and hour in seven to ten days following their laboratory period. Examination of final grades reveals no evidence indicating that laboratory sections are significantly different. The randomly selected students are expected to be statistically similar, and performance between each group should only change due to instructional interventions. The large class size ( $\approx 300$ ) allows us to have a sufficient number of students in each group to obtain reasonable statistical reliability. The data reported here are therefore not longitudinal data but randomized between-student data.

The data presented in this Brief Report were taken in 2008, during the second of a three-quarter traditional calculus-based introductory physics sequence taken primarily by first-year nonhonor engineering students at Ohio State University, a large public research university. The course covers electricity and magnetism and follows a typical syllabus. The course was an “off-sequence” class with approximately 300 students enrolled in two lecture sections. The class used WebAssign<sup>14</sup> for homework.

As an additional part of their homework grade, during the quarter each student was required to visit our research laboratory for a single 1 h session called the “flexible homework” assignment. During the session, each student took a variety of paper- and computer-based assessments on topics drawn from the syllabus. All students were told that they would not be graded on performance but rather would be given full credit for participation (equal to one homework assignment) and that they were to answer the best they could. Anecdotally, from observations of student work and postsession debriefings, we found that nearly all students made a good faith effort to answer the questions. About 95% of students in the class participated, and there is no correlation (or other discernible pattern) between final grade in class and date or day of week of flexible homework session. Because such a large percentage participated, any selection effects related to sampling only a certain population of students is greatly reduced.

In this Brief Report, we report on two question sets related to electrostatics (we also take data on other E&M and mechanics topics). Each set took about 10 min to complete. Our data collection started on quarter day 4 (QD 4), near the end of the first week, and ended on QD 45, the end of the ninth week. We did not collect any data in the tenth (last) week of the quarter.

Set 1 was composed of nine questions about electric fields and potentials for different arrangements of point charges. Each question presented one or two point charges, positive or negative, in several configurations. Students were asked to characterize the electric field and the electric potential as either rightwards, leftwards, rightward or leftward and positive, negative, or zero at one point of interest. They were also asked to compare the electric fields and electric potentials at two different points (some examples below). In all, 221 students completed this task and received final grades; 7 students were removed because they completed the task but did not receive a final grade (average score for removed students was 0.57, compared to 0.58 for included students).

Set 2 was composed of three questions from the conceptual survey of electricity and magnetism<sup>15</sup> (CSEM). From historical pre-post data for this population, we selected CSEM questions which had large absolute gains and which were thematically related. Set 2 contained (in order) CSEM questions 10 ( $g=0.19$ ), 15 ( $g=0.22$ ), and 13 ( $g=0.20$ ), which are about the electric force on a charge in a uniform field, electric field lines and negative charges, and the electric field inside a conductor, respectively. In all, 234 students completed set 2 with a final grade (11 students removed because they did not receive a final grade; average score for removed students was 0.33, compared to 0.41 for included students), 175 of whom also completed set 1.

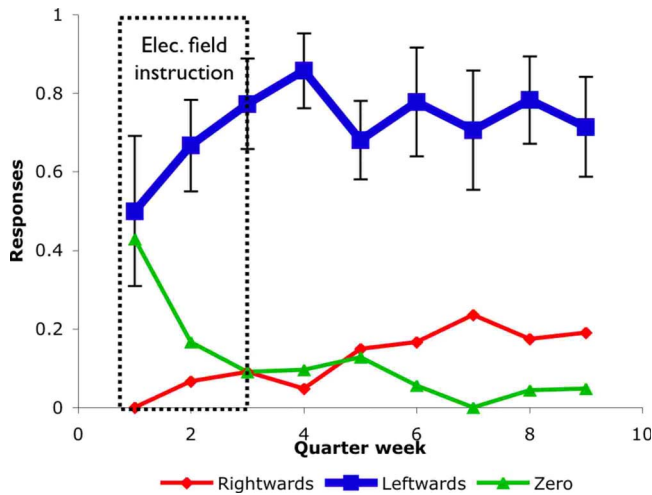


FIG. 2. (Color online) Responses to a question about the electric field between two charges of opposite signs. The correct answer (blue and bold) increases when the topic is taught and stays high for the remainder of the quarter. Some minor responses have been omitted.

#### IV. RESPONSE CURVES

As might be expected for thematically related questions, scores on similar questions within a set were similar. For set 1, the overall average score was  $72\% \pm 20\%$  on field questions and  $46\% \pm 15\%$  on potential questions. Set 2 had an overall average score of  $39\%$  and about  $10\%$  of students got all three correct. In accordance with historical data on questions in set 2, the average gain is about  $0.2$ . We report “response curves:” the weekly and/or daily answer choice averages as a function of time in the quarter. While some response curves on set 1 show little change over the quarter, many curves show interesting structure not measured by pre-testing and post-testing. In this Brief Report, we will focus on responses to four questions from set 1 and one question from set 2.

##### A. Learning

Figure 2 shows the response to a question about the electric field halfway between the two charges of opposite sign with the positive charge on the right. The correct “vectorlike” answer, in which the field points toward the negative charge (leftwards), and the leading “scalarlike” distractor, in which the field is zero, start the quarter with about equal proportions of responses. When the course covers electric fields, the correct answer increases to about  $70\%$ , where it remains for the rest of the quarter. As the course moves on from simple electrostatics, another response climbs, ending the quarter at about  $20\%$ . This response correctly attributes a vector nature to the field but reverses the sign convention. In many respects, this represents a pedagogically near-ideal curve: when the students learn a topic, it stays learned (with a possible confusion about sign).

##### B. Peak and decay

In contrast, Fig. 3 is more typical. The question concerns the electric potential halfway between two equal positive

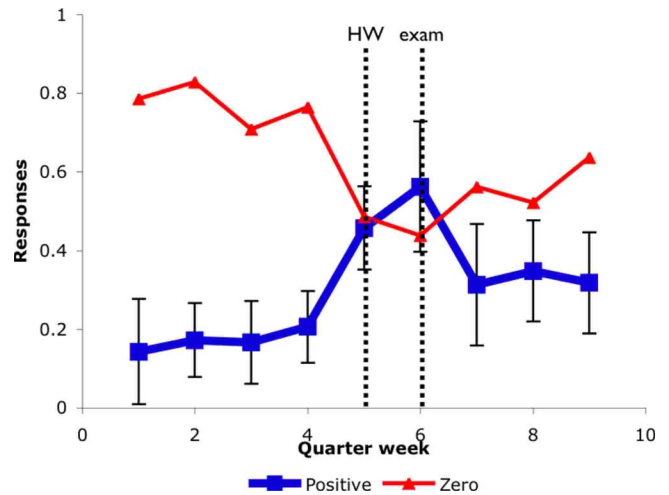


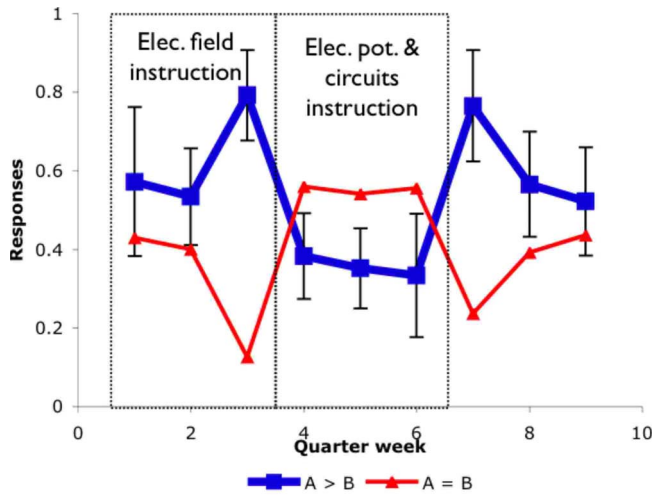
FIG. 3. (Color online) Responses to a question about the potential between two positive charges. The correct answer (blue and bold) increases in the middle of the quarter but decays afterward. Some minor responses have been omitted.

point charges. The correct scalarlike answer, in which the potential is positive, starts at about  $15\%$ , and the leading vectorlike distractor, in which the potential is zero, starts at about  $80\%$ . In weeks 5 and 6 of the quarter, the correct answer increases substantially, only to fall again in week 7. Pretesting and post-testing would not detect the peak because it is followed by the decay, and the decay is consistent with general patterns in memory research.

In the course, potential is covered starting on QD 15 (the end of the third week) and the next major topic (circuits) starts on QD 21 (the beginning of the fifth week), so it seems unlikely that lecture is the cause of this peak. The laboratory related to this material occurred in the third week of the quarter, so it seems an unlikely cause for the peak in weeks 5 and 6.

Four homework problems relevant to this problem were due on QD 23 at 8 a.m., the first day for which we collected data in the fifth week. In particular, one problem asks students to find the potential a distance  $d/2$  perpendicularly from the midpoint of the axis of two positive charges separated by distance  $d$ —a problem which is nearly the same as (but slightly more difficult than) the problem on this assessment. WebAssign gives students rapid feedback until they obtain the correct answer. Homework submission patterns suggest that most students work on problems within a few minutes of submission.<sup>16</sup> Thus it seems like the increase from week 4 to week 5 is due to this homework assignment.

The correct response on this question continues to climb in week 6, even though lecture, laboratory, and homework do not concern this topic in that week. The single midterm exam was on QD 29, the fourth day of week 6. It is possible that exam-related review is the major source of the continued peak in week 6. Graduate teaching assistants led an unofficial review session on QD 28, and lecture on QD 28 was devoted to exam review. In week 7, after the exam, the response decreases.



At which point is the Electric Field greater?

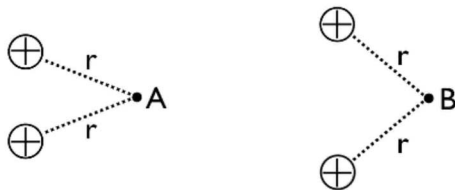


FIG. 4. (Color online) Prompt and responses to a question comparing the electric fields near two pairs of positive charges. The correct answer (blue and bold) decreases dramatically when potential is taught, then recovers. The third response, that  $B > A$ , was chosen in  $<7\%$  of all responses.

C. Interference

Figure 4 shows two responses to an electric field question about two pairs of charges (whose arrangements are shown in the figure).<sup>17</sup> The correct answer, in which the field is greater at point A than point B, can be seen from superposition and the vector nature of the electric field. The leading distractor is consistent with superposition of a scalar interpretation.<sup>18</sup> In week 3, the correct answer peaks as students learn about the vector nature of the electric field. However, in weeks 4–6, when lecture and homework concern the electric potential (first by point charges and later in circuits), the correct answer drops precipitously, only to recover in week 7 when magnetic fields are introduced.

The midquarter drop could be related to interference between existing knowledge of the vector nature electric fields and new knowledge about the scalar nature of electric potential. While the class is focused on potentials, the scalarlike response dominates. When the class returns to fields, the vectorlike response recovers.

D. All in one

For set 2, the structure of the response curve is most dramatic on the question concerning the motion of a particle in a uniform electric field. Figure 5 shows the popular responses to this question. Initially, students choose a variety of responses. On QD 13 there is a sharp peak (100%) in the correct answer, “constant acceleration,” after which it decays

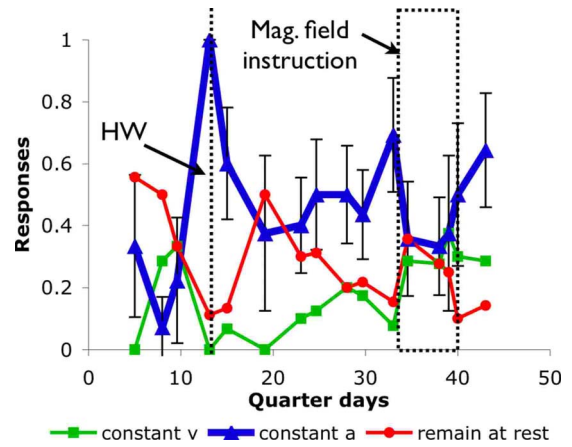


FIG. 5. (Color online) Responses by day on a question about the subsequent motion of a charge placed at rest in a uniform field. The correct answer (blue and bold) increases dramatically and falls rapidly. Later, during the unit on magnetic forces and fields, there may be an interference dip. Some minor responses have been omitted.

rapidly and stays constant except for a dip near QD 37. Concurrent with that dip, two incorrect answers increase: “constant velocity” and “remains at rest.”

The timing of the peak, about a week after this topic was covered in lecture, suggests that lecture is not a direct cause of the peak. The peak occurs on Wednesday; laboratories were open on Wednesday, Thursday, and Friday, so it is unlikely that laboratory is the cause. An investigation of homework sets reveals that a directly relevant homework problem (asking for the final velocity of a particle placed at rest in a uniform field) was due on QD 13 at 8 a.m., the day of the spike. In this homework problem, 80% of students who submitted the problem did so within 30 h of the due time and 60% within 12 h (60 students sampled; 46 submitted). Because the average grade on this question was 71% (for both sampled students and all students in the class), we can assume that most submitters answered the question correctly. It seems likely that this homework problem caused the sharp peak in correct responses.

The peak on this question is quite narrow, on the order of one day. Each data point represents 9–25 students (median 15) over one or two days; the peak represents ten students, all of which answered this particular question correctly. It seems unlikely that those responses are due to chance (the chance of ten correct responses on that day at random is about  $\approx 0.1\%$ , considering that about half of students answer correctly on nearby days). An alternative explanation is that only “good” students visited our laboratory that day. There are two reasons to refute this hypothesis. First, there is no correlation between final grade and day visiting our laboratory. Second, and more strongly, if those students were exceptional, we would expect a corresponding increase in their responses on other questions on the same instrument. We do not observe this increase, though responses generally track at all other points in the quarter. Thus, it is unlikely that the increase is caused by a sampling error.

The dip at QD 37 is much less pronounced than the spike, and it is broader. On QD 33, magnetic fields and the right-hand rule were introduced in the lecture. The first homework

for moving charges in magnetic fields was due on QD 38 and contained two right-hand rule questions. Interference between learning about magnetic forces and fields and prior knowledge about electric forces and fields could cause a dip near QD 38.

In accordance with the interference hypothesis, the response on the electrostatic task consistent with the behavior of charges in uniform magnetic fields increases during this time. A charge at rest in a magnetic field will remain at rest (red line in Fig. 5). This dip is not as pronounced as for the questions in task 1, and it occupies less time.

## V. CONCLUSION

We have presented evidence for three well-established and fundamental cognitive phenomena—learning, decay, and interference—occurring in the real world context of a traditional physics class. These were measured by repeatedly testing different and random groups of students throughout the duration of the course and by direct comparison to what was taught in the course at specific times. Each of the test questions was multiple choice and focused on a very specific concept, and for many questions there was little change in student performance throughout the course. However, for some questions such as those presented in this Brief Report, there were significant and rapid changes in student performance.

Learning was characterized by a rapid increase in performance. Interestingly, the increase in performance did not coincide with relevant lectures or laboratories for the questions reported here but increased directly after immediate-

feedback homework assignments closely related to the test questions. We also found very rapid decays in performance, on the order of days after the peak in performance. This is consistent with memory research, which documents exponential decay in memory; however the short time scale found here is somewhat unexpected.

Interference from learning one topic on the performance of another previously learned related topic was also observed. In particular, it was found that learning about electric potential significantly decreased the student's ability to recognize the vector nature of the electric field. After the interfering unit of potential and the course returned to vector fields, the interference between electric field and potential disappeared. We find that interference can also decay rapidly with time.

In summary, we find that student performance can have rapid and large peaks and decays on the time scale of days that would otherwise be missed by pretesting and post-testing. Such information provides important information about the effectiveness of particular instruction methods and indicates that memory decay and interference can play an active role in student performance.

## ACKNOWLEDGMENTS

We thank Jeffrey Asch, Rebecca Rosenblatt, Thomas Scaife, and the others in the Physics Education Research Group and the Cognitive Development Laboratory at the Ohio State University for their insight on this project. This research is partially supported by a grant from the Institute of Education Sciences, U.S. Department of Education (Grant No. R305H050125).

- 
- <sup>1</sup>R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- <sup>2</sup>D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- <sup>3</sup>D. Hestenes and M. Wells, A mechanics baseline test, *Phys. Teach.* **30**, 159 (1992).
- <sup>4</sup>E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- <sup>5</sup>R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- <sup>6</sup>R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- <sup>7</sup>L. Ding, N. W. Reay, A. Lee, and L. Bao, Effects of testing conditions on conceptual survey results, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010112 (2008).
- <sup>8</sup>G. B. Semb, J. A. Ellis, and J. Araujo, Long-term memory for knowledge learned in school, *J. Educ. Psychol.* **85**, 305 (1993).
- <sup>9</sup>R. A. Rescorla and A. R. Wagner, in *Classical Conditioning II*, edited by A. H. Black and W. F. Prokasy (Appleton-Century-Crofts, New York, 1972), pp. 64–99.
- <sup>10</sup>C. E. B. Hermann Ebbinghaus, *Henry Alford Ruger, Memory* (Columbia University, New York, 1913).
- <sup>11</sup>M. E. Bouton, Context, time, and memory retrieval in the interference paradigms of Pavlovian learning, *Psychol. Bull.* **114**, 80 (1993).
- <sup>12</sup>L. Postman, Critical issues in interference theory, *Mem. Cognit.* **1**, 19 (1973).
- <sup>13</sup>Here, we will not consider in detail the nature or frequency of the interventions during this period rather only assume that it occurs a number of times during the intervention period.
- <sup>14</sup>J. C. Dutton, *WebAssign: A Better Homework Delivery Tool* (The Technology Source, 2001).
- <sup>15</sup>D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. V. Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- <sup>16</sup>For example, if students were to work all problems ahead of time, we might expect to see all questions submitted within a few minutes of each other. Instead, many students submit questions spaced out over the course of several hours, indicating that they submit responses as they work.
- <sup>17</sup>To save space in this Brief Report, the figure depicts points *A* and *B* closer to each other than in students' task. While it is possible that some students could have thought that the fields at points *A* and *B* were due to all four charges, we do not have any evidence (from debriefing interviews or otherwise) that students misinterpreted the task in this manner.
- <sup>18</sup>There is a companion question which concerns the electric potential for the two pairs of charges, and it exhibits similar behavior to the field question.