

Applying module analysis to the Conceptual Survey of Electricity and Magnetism

Christopher Wheatley,¹ James Wells², Rachel Henderson,^{3,4} and John Stewart^{1,*}

¹West Virginia University, Department of Physics and Astronomy,
Morgantown, West Virginia 26506, USA

²University of Connecticut, Department of Physics, Storrs, Connecticut 06269, USA

³Michigan State University, Department of Physics and Astronomy, East Lansing, Michigan 48824, USA

⁴CREATE for STEM Institute, Michigan State University, East Lansing, Michigan 48824, USA



(Received 8 September 2020; accepted 5 January 2021; published 19 January 2021)

The Conceptual Survey of Electricity and Magnetism (CSEM) is a widely used multiple-choice instrument measuring a student's conceptual understanding of electricity and magnetism. This study applied modified module analysis (MMA) and modified module analysis—partial (MMA-P), network analytic methods that identify groups of correlated responses, to CSEM data from two institutions ($N = 2538$ and 3595). In module analysis, groups of correlated responses are called “communities.” As in previous applications of MMA and MMA-P to mechanics conceptual inventories, a number of communities related to physics concepts and some communities related to the structure of blocked items in the inventory were identified. An item block is a set of items all referring to each other or to a common stem. Many blocked communities involved responses where the response to the later item would be correct if the response to the earlier item was correct. This suggests a modified scoring rubric for the CSEM is needed to account for these connections between items. A modified scoring rubric is proposed; however, the modified overall average scores changed by less than 1%. The communities of incorrect responses to the CSEM related to physical concepts had varied explanations. These explanations ranged from seemingly straightforward errors (the electric field pointing to higher potential or reversing the right-hand rule), to misconceptions about Newton's 2nd and 3rd laws carried over from mechanics, to naive reasoning conflating general topics in electricity and magnetism. The identification of incorrect communities allowed the computation of misconception scores showing how prevalent the misconceptions were in the classes studied.

DOI: [10.1103/PhysRevPhysEducRes.17.010102](https://doi.org/10.1103/PhysRevPhysEducRes.17.010102)

I. INTRODUCTION

Students' conceptual understanding of physics and coherently applied errors in that understanding, misconceptions, have long been important research areas within physics education research (PER). This research has been fostered by the introduction of multiple-choice conceptual instruments such as the Force Concept Inventory (FCI) [1], the Force and Motion Conceptual Evaluation (FMCE) [2], the Conceptual Survey of Electricity and Magnetism (CSEM) [3], and the Brief Electricity and Magnetism Assessment (BEMA) [4]. Recently, network analytic techniques, called modified module analysis, have been applied to the FCI and FMCE [5–8] and have identified common student incorrect answering patterns as well as potential flaws in the instruments.

A. Research questions

The current study applies modified module analysis (MMA) to the CSEM to investigate correlated patterns of student responses to the instrument. MMA forms a network where the responses to the items in a multiple-choice instrument are the nodes and the edges represent the correlation between the responses. Network analysis identifies “communities” within the network of responses that are often selected together by a student. Two versions of MMA have been used to explore conceptual physics instruments, one using the correlation matrix (MMA) [6], the other the partial correlation matrix (MMA-P) [8]. The current work applied both versions to two large samples of CSEM responses from two different institutions. This study explored the following research questions:

RQ1 What community structure is identified by network analysis of the CSEM? How are the communities associated with previously identified features of the instrument?

RQ2 Does the community structure of the CSEM have communities related to Newtonian mechanics? If so, how do these communities compare to the communities identified in the FCI or the FMCE?

*jstewart1@mail.wvu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

RQ3 How do the communities identified by the two versions of module analysis, MMA and MMA-P, compare? How do the communities identified at different institutions compare?

This work identified a rich ecology of diverse types of incorrect reasoning, much broader than that identified in the FCI or FMCE. It used the community structure identified to calculate scores representing the relative strength of types of incorrect reasoning and offered a modified scoring rubric for the CSEM which corrects for the relations found between items.

B. Prior studies

This work makes extensive use of the results of four prior studies which will be referenced as study 1 to study 4.

1. Study 1

In study 1, Maloney *et al.* [3] introduced the CSEM, provided a classification of items in the instrument, and discussed common errors made by students both pre- and postinstruction. The CSEM was created by combining two prior surveys about electricity and magnetism by Hieggelke and O’Kuma [9]. The CSEM is an instrument with 32 items that measures a student’s knowledge of concepts in electricity and magnetism. The instrument includes questions about topics commonly covered in introductory electricity and magnetism courses, such as conductors and insulators, Coulomb’s law, superposition, electric fields, magnetic fields, and magnetic induction. The construction of the instrument was informed by a number of prior studies discussed in Sec. 1D. This study uses the version available at PhysPort [10].

Study 1 provided a general classification of the items in the CSEM: charge distribution on conductors or insulators (items 1, 2, 13), Coulomb’s force law (items 3, 4, 5), electric force and field superposition (items 6, 8, 9), force caused by an electric field (items 10, 11, 12, 15, 19, 20), work, electric potential, field, and force (items 11, 16, 17, 18, 19, 20), induced charge and electric field (items 13, 14), magnetic force (items 21, 22, 25, 27, 31), magnetic field caused by a current (items 23, 24, 26, 28), magnetic field superposition (items 23, 28), Faraday’s law (items 29, 30, 31, 32), and Newton’s 3rd law (items 4, 5, 7, 24).

Study 1 also discussed common errors made by students; these errors were not referred to as misconceptions, but were often made on multiple items suggesting coherently applied incorrect knowledge. Students confuse the behavior of conductors and insulators (items 1 and 2). Students do not fully understand the shielding of the electric field by conductors (items 13 and 14). Responses to item 14 also show a failure to understand Newton’s 3rd law; this misunderstanding is also detected by items 4, 7, and 24. Students apply the larger object exerts more force misconception [11] on item 4; this misconception was also identified within the FCI in study 3 and the FMCE [7].

On items 8 and 9, response D, students misunderstand how the addition of another charge affects the field. In the current study, response D to item 8 is abbreviated as response 8D. Students confuse the behavior of electric and magnetic fields in responses 23B, 23C, and 26B. Response 10B represents the force proportional to velocity misconception [11] also detected in the FMCE [7]. On items 19 and 20, students confuse the relation of changes in electric potential to the direction of the electric field.

2. Study 2

Prior network analytic studies have made extensive use of constrained multidimensional item response theory (MIRT) models of the correct physical reasoning needed to solve the items in the FCI [12] and FMCE [13]. This work utilized a similar study of the CSEM [14] which is referenced as study 2 in this work. All three studies identified the practice of “blocking” items into item groups as a source of correlations within the responses to the instrument. A group of items is blocked if all items in the group refer to a common stem describing the physical system or if one item explicitly refers to a prior item in the block. The CSEM contains 3 item blocks {3, 4, 5}, {10, 11}, and {17, 18, 19}. In study 2, items 4, 5, and 11 were eliminated from the analysis and only the first item in the block was retained because the answers to the latter items in the group directly depended on the earlier items. Items 18 and 19 were retained because it was felt that all items in the block could be answered independently. Study 2 also identified 3 groups of isomorphic items {6, 8}, {16, 17}, and {21, 27}. Isomorphic items all require the same solution process. In the prior network analytic studies of the FCI and the FMCE, responses to isomorphic items have often been detected both in the same correct communities and the same incorrect communities.

C. Prior network analytic studies

Module analysis is a technique for analyzing multiple-choice instruments using network analysis [5–8]. A network is a representation of a system as a set of nodes connected by edges to form a graph. Information related to the objects or concepts that the nodes and edges represent can be included in the network [15]. Module analysis connects responses selected together by students, then uses a community detection algorithm (CDA) to partition the network into communities. These communities, also called modules, are then examined to determine what the set of items reveals about student thinking.

Initially, Brewe *et al.* applied module analysis for multiple choice responses (MAMCR) to incorrect responses of the FCI post-test results of 143 first year physics majors at a university in Denmark [5]. Incorrect responses were made into a network where the edges represented the number of times two items were chosen together. If the correct responses were included, they

formed a single community that obscured other structure. They found nine communities, but only three of them could be assigned an unambiguous interpretation of the underlying incorrect idea.

The work of Brewe *et al.* inspired a series of studies applying modifications of their original algorithm to conceptual inventory data.

1. Study 3

In study 3, Wells *et al.* created the modified module analysis technique to study the FCI [6]. MMA uses the correlation between the responses to form edges in the network, which scales to larger networks better than MAMCR. The communities that result from MMA tend to only include a pair or a small number of responses, which allows the underlying idea that may lead a student to select those responses to be identified. MMA was applied to 4509 pretest and 4716 post-test FCI records of students in an introductory calculus-based physics class. Some of the structure found by MMA was due to blocked items within the instrument. Excluding these blocked items, there were five communities on the post-test that represented Newton's 3rd law misconceptions, the motion implies active forces misconception, the motion implies active forces for the centrifugal force misconception, the circular impetus misconception, and the largest force determines motion misconception, as described in Hestenes and Jackson's taxonomy of misconceptions measured by the FCI [11].

In a later study, MMA was also applied to the FMCE pretest ($N = 3956$) and post-test records ($N = 3719$) from students in an introductory calculus-based physics course [7]. The FMCE is a highly blocked instrument with seven main item groups (the four energy questions added in 1999 [10] were not analyzed), some of which test the same physical principle in a different representation, either a written description or as graphs. Two of the five post-test modules represented the velocity proportional to applied force misconception, two represented the velocity-acceleration indiscriminated misconception, and one module applied both the greater mass implies greater force and the most active agent produces largest force misconception from the Hestenes and Jackson's taxonomy [11].

2. Study 4

In study 4, in order to include both correct and incorrect responses in a module analysis, modified module analysis using partial correlations (MMA-P) was developed by Yang *et al.* [8]. In MMA-P, the partial correlation matrix is used to connect each pair of responses, controlling for the effect of total score. MMA-P can identify communities that include only correct responses, only incorrect responses, or both correct and incorrect responses. As such, a richer set of communities are found with MMA-P than with MMA. MMA-P was applied to the same sample of FCI responses as in study 3. The completely incorrect communities were

very similar to those identified by MMA. The completely correct communities generally involved blocked items or were identified as isomorphic by MIRT [12]. The mixed communities suggested some items in the FCI were not functioning correctly.

D. Studies informing the construction of the CSEM

Mechanics instruments such as the FCI were written after a great deal of research had been performed on students' conceptual understanding of mechanics. Substantially less research had been performed on electricity and magnetism before the CSEM was published. This section summarizes some of the work that informed the construction of the CSEM.

Maloney [16] performed a study using activities that tested students' conception of magnetic poles post-instruction in a general physics course and found that the majority of students had an alternate conception based around the idea that "magnetic poles are charged." Guruswamy *et al.* [17] studied student understanding of the transfer of charge between conductors through a small set of questions about simple charge transfer experiments. The vast majority of students tested, from 8th grade to physics majors in senior-level university physics courses, could not correctly explain or predict what happens when charged conductors come in contact with each other. Törnkvist *et al.* [18] investigated the understanding of electric fields by introductory college students. They concluded that the majority of students personify field lines as isolated entities in real space, rather than a set of curves that represent mathematical properties of space. Galili [19] studied high school student difficulties with the field concept in electricity and magnetism. Students often regressed in their understanding of mechanics concepts that were previously understood when learning about the concept of fields in electricity and magnetism. Studies of student reasoning about and understanding of the superposition of electric fields have shown that many students struggle with causality in electricity and magnetism [20,21]. Most notably, some students do not recognize the existence of a field unless there is some motion caused by the field.

E. Prior studies of the CSEM

While not as thoroughly studied as the FCI, multiple studies have used the CSEM to explore student conceptual thinking about electricity and magnetism.

Planinic [22] compared Croatian students to American students in a study that introduced six overarching conceptual areas measured by the CSEM. In order to produce groups large enough for analysis, Planinic qualitatively regrouped the eleven concepts reported in study 1 into six categories. These areas include electric charge and force (items 1, 2, 3, 5, 6, and 8), Newton's laws (items 4, 7, 10, 24, and 27), electric field and electric force (items 9, 12, 13, 14, 15), electric potential and energy (items 11, 16, 17, 18,

19, 20), magnetic field and magnetic force (items 21, 22, 23, 25, 26, and 28), and induction (items 29, 30, 31, 32). The difficulty of the individual items in each group was very similar for the two populations [22].

Performance differences between men and women on the CSEM have also been explored. A difference in CSEM test performance by gender was measured by Kreutzer and Boudreaux [23] and was greatly reduced by pedagogical changes. Gender differences on the CSEM were also examined by Kohl and Kuo [24] through a transition to studio physics. Studio physics is a model of instruction where students take an active role in learning by doing hands-on activities and group work during instruction rather than in separate labs. They found the gap in normalized gain was reduced by this transition [25]. Henderson *et al.* examined gender differences in performance on the CSEM and compared these to gender differences in other multiple-choice problems in a university physics class [26]. A 5% difference in CSEM post-test scores was measured; however, the difference in conceptual in-semester test questions, which did not include CSEM questions, was only 3% with no difference observed in quantitative test questions. For a more complete summary of gender differences in conceptual understanding of physics see Madsen, McKagan, and Sayre [27].

Other studies have analyzed a subset of items in the CSEM. Leppävirta examined Newton's 3rd law using items 4, 5, 7, and 24 and showed that the number of students with an incorrect model of Newton's 3rd law decreased from 20% to 10% from pretest to post-test [28].

Changes from pretest to post-test were also explored by Meltzer using items 18 and 20 to investigate how electric field concepts intersect with potential concepts [29]. Meltzer reported that students' conflation of electric field magnitude with potential slightly increased post-instruction. Study 1 also reported the conflation of electric and magnetic fields. All other incorrect response pairs between items 18 and 20 decreased postinstruction, while the correct response pairs significantly increased.

Karim *et al.* [30] used the CSEM to study the degree to which graduate teaching assistants (TA) could predict introductory physics students' alternate conceptions in electricity and magnetism. TAs were told to choose the response that they thought would be the most chosen incorrect response by the students in the introductory course. The TAs were likely to choose responses that included both correct and incorrect concepts, but their choices did not correspond to the most frequent incorrect responses by students in the course.

F. Theories of knowledge

Physics education research has investigated common student difficulties with conceptual physics since its inception. These difficulties have often been conceptualized as "misconceptions" or "alternate conceptions/hypotheses."

Early work [31–33] analyzed common alternate views of force and acceleration. Halloun and Hestenes [34,35] extended these works by developing a taxonomy of "common sense concepts" representing incorrect reasoning about Newtonian mechanics. The FCI was developed partially with the goal of measuring these incorrect reasoning patterns [1]. The authors of the FCI provide a detailed description of the misconceptions measured by the instrument. This description was refined by Hestenes and Jackson [11] to produce the taxonomy applied in study 3, 4, and Wells *et al.* [7]. We used this taxonomy in the present work.

In the current work, network analysis identified two types of Newtonian mechanics misconceptions described by Hestenes *et al.* [1]. Within the "active force" group of misconceptions, the "velocity proportional to applied force" misconception showed that Newton's 2nd law was not well understood. A student applying this misconception reasons that the velocity of a particle in motion will be equal or proportional to the force applied to the object. This misconception was also identified in the FMCE by Wells *et al.* [7] using MMA.

Two misconceptions involving a misunderstanding of Newton's 3rd law were identified as forming the "action/reaction pairs" group of misconception by Hestenes *et al.* [1]: "greater mass implies greater force" and "most active agent produces greatest force." Students applying the greater mass implies greater force misconception reason that the larger or heavier object exerts more force than the smaller or lighter object. Students applying the most active agent produces greatest force misconception assume an active object produces more force than an inactive object; for example, a small car pushing a large truck exerts more force on the truck than the truck exerts on the car. Module analysis identified both misconceptions in the FCI (study 3) and the FMCE [7]. In both studies, responses demonstrating both action/reaction pairs misconceptions were found in the same community.

The misconception model of incorrect reasoning was important in the development of many conceptual physics instruments, particularly the FCI and FMCE. Other models of incorrect thinking have also been used to explain student reasoning in physics. Two of the most important models are ontological categories [36–38] and knowledge in pieces [39,40]. The ontological categories theory explains incorrect student reasoning as the misclassification of some quantity [36–38], for example the misclassification of force as a substance. If that substance can be used up, then one would predict that an object in motion after the application of a force would come to a stop as the force was used up. The knowledge-in-pieces framework suggests that student incorrect and correct reasoning is composed of small pieces of reasoning that are activated either singly or collectively to address a problem [39,40]. Many researchers have explored variations of this model, conceptualizing the

reasoning fragments as phenomenological primitives (p primes) [39,40], facets of knowledge [41], or resources [42–44].

Scherr provides a definition that contrasts the knowledge-in-pieces and misconceptions views which we adopted in this work [45]. The misconception view is “a model of student thinking in which student ideas are imagined to be determinant, coherent, context-independent, stable, and rigid,” [45] while knowledge-in-pieces models student conceptions “as being at least potentially truth-indeterminate, independent of one another, context-dependent, fluctuating, and pliable” [45].

The examples of misconceptions given above are familiar to many physics instructors and researchers; p primes may be less familiar. In a 2018 introduction to the knowledge-in-pieces framework, diSessa provided the following examples of p primes: *Example p primes are (roughly described): increased effort begets greater results; the world is full of competing influences for which the greater “gets its way,” even if accidental or natural “balance” sometimes exists; the shape of a situation determines the shape of action within it (e.g., orbits around square planets are recognizably square)* [46]. As such misconceptions represent the application of alternate forms of reasoning, which could have been part of a coherent theory had the universe been different, while p primes are more ephemeral representing often intuitive pieces of reasoning applied differently across diverse contexts which have little physical relation.

For this work, we identify communities of incorrect responses as misconceptions if the student seems to apply consistent (incorrect) reasoning to physical situations that would be viewed as similar by an expert. For example, if a community is detected where the students consistently reason that the larger object in a collision exerts the greater force they are applying the “greater mass implies greater force” misconception. However, communities where students seem to apply similar reasoning to situations that would not be viewed as related by experts, for example, reasoning the larger object exerts a larger force in a collision and larger equipotential spacing implies stronger electric field, would be identified as an example of the application of a p prim such as “increased effort begets greater results.” MMA-P, which can identify communities containing both correct and incorrect responses, may offer stronger evidence by finding communities applying the same reasoning to reach both correct and incorrect conclusions. For example, the “increased effort begets greater results” might also be used to reason that greater potential difference implies greater current.

In the work introducing the FCI, Hestenes *et al.* [1] differentiated naive conceptions from misconceptions and reported that some misconceptions were weakly held and some were strongly held [1]. They separate naive conceptions of kinematics where various combinations of

position, velocity, acceleration, and force are not differentiated in student thinking from more robust incorrect models similar to medieval theories of motion. While not specifically named in Hestenes *et al.*, the term naive conception is introduced in the taxonomy of Hestenes and Jackson [11]. In the current work, we use naive conception for student reasoning that convolves multiple physics concepts and misconception for more robust modes of incorrect reasoning.

The FCI and FMCE were developed within the misconception framework and network analysis largely supported this framework identifying only one community seeming to correspond better to a p prim than to a misconception. The CSEM was not developed from a robust framework of misconceptions, but rather responses were taken from common open-response answers to the items. No robust taxonomy of misconceptions of electricity and magnetism similar to Hestenes and Jackson’s taxonomy of misconceptions of mechanics [11] has been published. This offers the possibility that a broader set of structures not identifiable as misconceptions may be identified using MMA or MMA-P.

MMA and MMA-P are quantitative methods that identify consistently selected responses to a multiple-choice instrument. They cannot identify the theory of knowledge which best represents the student thinking that generated the consistent responses; additional qualitative research would be needed to make this determination.

II. METHODS

A. Sample

This study was performed on samples from two US institutions.

1. Sample 1

Sample 1 was collected from spring 2003 until spring 2012 at a southern land-grant university with a total enrollment of about 25 000 students. The demographics of the undergraduates at the university were 76% White, 9% Hispanic/Latinx, 4% African American/Black, 4% two or more races, 2% Asian with other groups composing no more than 1% [47]. International students, students who are citizens of countries other than the U.S., were 3% of the sample. The overall undergraduate population had ACT scores ranging from 23–29 (25th to 75th percentile) [47]. In the course studied, 77% of the students were men.

The CSEM post-test was given in the introductory calculus-based electricity and magnetism courses serving scientists and engineers. Only students with complete post-test responses were retained for the study ($N = 2538$). The CSEM was given as a quiz after instruction and the student’s scores were recorded for part of their course grade. The same professor taught the course for the total period studied. The course implemented a number of

interactive engagement instructional practices in both the lecture and laboratory.

2. Sample 2

Sample 2 was collected from Fall 2015 to Spring 2019 at an eastern land-grant university with a total enrollment of about 30 000 students. The overall undergraduate demographics were 80% White, 4% African American/Black, 4% Hispanic/Latinx, 4% two or more races, 2% Asian with other groups composed of less than 1% [47]. International students were 6% of the sample. The overall undergraduate population had ACT scores ranging from 21–26 (25th to 75th percentile) [47]. In the course studied, 80% of the students were men.

The CSEM post-test was given in the introductory calculus-based electricity and magnetism courses serving scientists and engineers. Only students with complete post-test responses were retained ($N = 3595$). The CSEM was given as a quiz postinstruction and graded for a small amount of course credit. The course was managed by a single lead instructor in the time studied who taught the majority of the lecture sections. Interactive engagement methods were applied in both the lecture and laboratory.

B. Correlation and partial correlation

This work uses both the correlation, r_{XY} between two dichotomously coded responses X and Y to the CSEM, and the partial correlation, $r_{XY|Z}$ controlling for overall CSEM score Z . The correlation, r_{XY} between two responses X and Y , is commonly used in PER and is defined by

$$r_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

where μ_i is the mean of variable i , σ_i is the standard deviation, and $E[X]$ is the expectation value.

The partial correlation is less familiar. Two responses to an instrument may be correlated because they are only answered correctly by the highest scoring students; these responses are correlated through total test score. A partial correlation controls for the effect of the third variable Z on the correlation between X and Y by removing the contribution of the common (mathematical) dependence of X and Y on Z . The definition of the partial correlation between X and Y correcting for Z is given by

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{1 - r_{XZ}^2}\sqrt{1 - r_{YZ}^2}}. \quad (2)$$

Qualitatively, partial correlation can be understood in terms of the linear regression of continuous variables. If both X and Y are related to a third variable Z , then the effect of Z on either X or Y can be controlled for using linear regression. If X is the dependent variable in a regression, with Z as the independent variable, then the residuals of the

regression contain the part of X not explained by its relation with Z . The partial correlation is the correlation between these residuals and the residuals of a similar regression with Y as the independent variable.

While this study uses the p value of the correlation to remove edges in the next section and therefore is susceptible to distributional assumptions; in this case with the large sample size and retaining only correlation values above 0.15 (the threshold for small effect size is 0.1 using Cohen's criteria [48]), the restriction to p values less than 0.05 removed no edges. All p values meeting the $r > 0.15$ threshold were extremely small.

C. Modified module analysis

Modified module analysis and modified module analysis—partial, as described in study 3 and 4, were applied to the CSEM. These methods are described in greater detail in these prior studies. A general summary follows.

Module analysis begins by forming a network of the responses to a multiple-choice instrument. Each response forms a node in the network. For example, response D to item 3 becomes node 3D. MMA and MMA-P differ in the way they construct the edges connecting the nodes in the network. In MMA, an edge connects two nodes if the correlation between the two responses r is larger than some threshold; $r > 0.2$ in study 3. Only incorrect responses are analyzed in MMA. In general, correct answers to a well constructed instrument are strongly positively correlated; if the correct answers are included in MMA, they form a tightly connected community which obscures other structures. In MMA-P, an edge connects two nodes if the partial correlation $r_{XY|S}$ controlling for total CSEM score S exceeds some threshold; $r_{XY|S} > 0.2$ in study 4. This modification allows the inclusion of correct responses in the MMA-P networks. The application of techniques to reduce the complexity of a network, such as applying the threshold, is called sparsification in network analysis. In MMA and MMA-P, nodes selected by fewer than 30 students were removed as statistically unreliable. An additional effect of the threshold was to remove the large negative correlations between different responses to the same item and to ensure that each edge retained in the network represented a significant correlation. Each edge was checked for significance at $p > 0.05$ after a Bonferroni correction was applied; at the sample size of this study and the thresholds used, all correlations were significant.

Module analysis then applies a community detection algorithm (CDA) to the network. The fast-greedy [49] CDA was applied in study 3 and 4 to identify communities within the network. A community is a set of nodes more closely related to each other than to nodes outside of the community. The fast-greedy algorithm is one of many community detection algorithms. Study 3 reported that most CDAs yielded similar outcomes for the correlation network. In order to account for random fluctuations, the data were

bootstrapped with 1000 replications and the CDA was applied to each replication. Bootstrapping is a statistical technique which samples a dataset with replacement to use the distribution of the data itself to draw statistical conclusions. The igraph package [50] and the boot package [51] in R were used for the CDA and for bootstrapping respectively. The network was divided into communities 1000 times sampling the data with replacement. The percentage of times that any pair of nodes appeared in the same community was calculated; this quantity was

called the community fraction C . Communities that were found in $C > 80\%$ of the bootstrapped samples were analyzed in this study as was done in study 3 and 4.

III. RESULTS

Modified module analysis and modified module analysis—partial were applied to the CSEM; Fig. 1 shows the communities detected with their respective correlation thresholds. Three of the four analyses were performed with

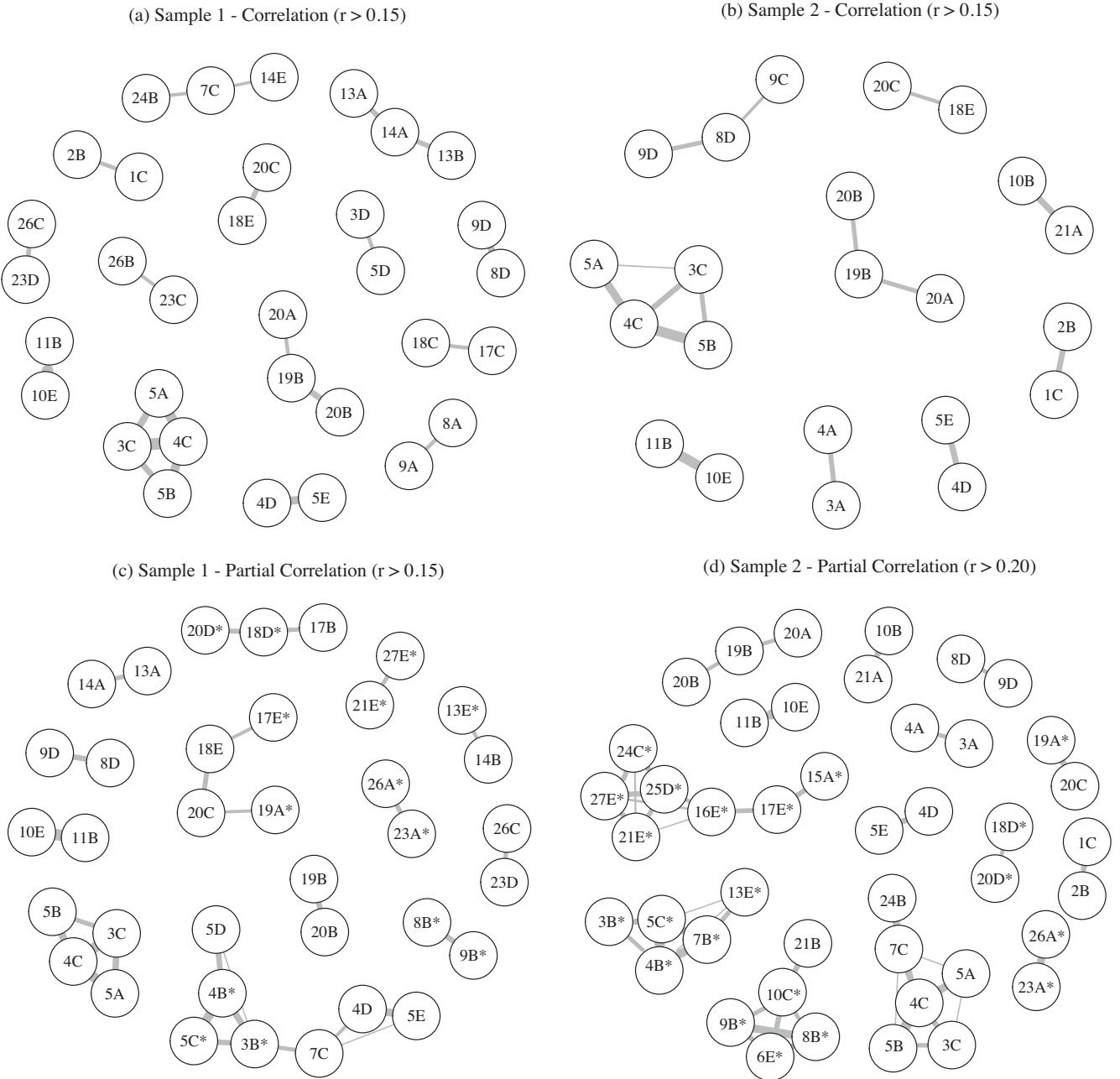


FIG. 1. Communities detected in the CSEM. (a) The correlation network for sample 1. (b) The correlation network for sample 2. (c) The partial correlation network for sample 1. (d) The partial correlation network for sample 2. The strength of the correlation or partial correlation is represented by the line thickness. Weaker correlation lines have been enhanced to be a minimum width for visibility.

$r > 0.15$ instead of the $r > 0.20$ threshold used in study 3 and 4. The threshold was adjusted to provide a fairly disconnected, but rich, set of communities. This is consistent with prior MMA and MMA-P studies and is necessary because the differences in the overall performance of each sample on the CSEM generates lower overall correlations in some samples and partial correlation values are generally lower than correlation values; therefore, the threshold at which the network becomes sparsely connected may differ. Adjusting the sparsification parameters is a common practice in network analysis. Study 3 presents a sequence of threshold values to demonstrate how the correlation network sparsifies with changing r threshold. Optimally setting the threshold should be an important direction in future research. The nodes of the networks that correspond to correct responses are labeled with an asterisk (*). The magnitude of the correlation or partial correlation between nodes in Fig. 1 is proportional to the line thickness.

The network graphs in Fig. 1 were rendered using the Fruchterman-Reingold network visualization in the *igraph* [50] package in R. The primary purpose of this visualization is to render a complicated network into a visually appealing form. This is done using a force-directed graph drawing algorithm where fictitious forces are introduced between nodes. Hooke's law like forces with strength proportional to edge weight are introduced between connected nodes and then repulsive Coulomb's law forces are introduced between all nodes. The system's equilibrium is then calculated from a random initial node placement. As such, the physical placement of communities or nodes within communities confers no useful information. The information presented is that which is important for network analysis: sets of closely interconnected nodes are drawn close together and communities without interconnection are drawn far apart. The general radius of the community is related to the overall average correlation between nodes in the community.

The total scores of the two samples were quite different. For sample 1, the CSEM post-test percentile score was 61.8 ± 0.3 ; for Sample 2, the post-test percentile score was 45.5 ± 0.3 . The scores are presented as mean $\pm$ standard error.

Table I provides a summary of the communities as well as a possible explanation of the common reasoning applied. A number of the communities identified were composed of items within item blocks. There is substantial evidence that the practice of blocking items produces correlations between the items that are not related to consistently applied reasoning [12,13]. These groups are labeled "blocked items" and are discussed separately. The communities are divided into three classes: communities composed entirely of incorrect responses, mixed communities composed of both correct responses and incorrect responses, and communities composed entirely of correct

responses. Some communities appear as independent communities in some samples, but as part of a larger community in others. These sub-communities are marked by $\times$ to show the continuity between communities more accurately. Examination of Fig. 1 shows that while many communities are completely connected, some are not. A community is completely connected if each node in the community has an edge to every other node in the community. In order to display the connectivity of individual communities in Table I, responses that are completely connected are shown in parenthesis, while responses that are only connected to each other are separated by dashes. For this notation, some particularly weak correlation lines were ignored so as to capture the most important structure of the network. As an example of this notation, consider the MMA-P post-test community in sample 2: 5A-(3C, 5B, 4C)-7C-24B. The group (3C, 5B, 4C) is completely connected; every node in the group has an edge to every other node. The pair 7C-24B is connected to each other but 24B is not connected to other nodes in the community. The node 7C is also connected only to node 4C in the group (3C, 5B, 4C) by an edge with a substantial correlation.

For this study, we grouped items in the CSEM into 4 broad topics: electrostatics (items 1 to 15), electric potential (items 16 to 20), magnetostatics (items 21 to 28), and magnetic induction (items 29 to 32). No communities were detected that involved the magnetic induction items. Examination of these items suggests that items 29, 30, and 32 should require related reasoning; failure to find responses to these items in the same community may indicate that the items are not functioning as intended. Generally, communities were formed within these broad topics with a few notable exceptions. Communities were identified that involved misconceptions of Newtonian mechanics that involved both electrostatic and magnetostatic items. Some students also answered questions about the electric and magnetic field in the same way producing communities representing the electric-magnetic field undiscriminated naive conception. Study 3 and 4 also identified both blocked items and isomorphic items as important in contributing to the formation of communities. The electric-magnetic field undiscriminated terminology was selected to mirror that of the misconception classifications of the FCI in Hestenes and Jackson's taxonomy [11].

A. Blocked items

The CSEM contains 3 item blocks {3,4,5}, {10,11}, and {17,18,19}. In the first two blocks, the answers to later items in the block are dependent on the answers to the earlier items in the block. Items 17, 18, and 19 can be answered independently, but could be correlated if a student misinterpreted the shared description of the physical system. The items in an item block are often found in the same community in Table I. Both study 3 and 4 also

TABLE I. Communities identified in CSEM responses. Communities labeled $\times$ are subcommunities of another community. Sample 1 is abbreviated S1; sample 2, S2. Responses in parentheses are completely connected. Responses separated by dashes are only connected to each other. Blocked items are marked consistent if later items apply correct reasoning to an incorrect earlier item.

Community	Correlation		Partial correlation		Misconception/Naive conception/Principle/ Explanation
	S1	S2	S1	S2	
Completely incorrect communities					
1C, 2B	⊗	⊗		⊗	Conductor and insulator misconceptions.
3A, 4A		⊗		⊗	Blocked items—consistent.
3D, 5D	⊗				Blocked items—(5D) $E \propto 1/r$.
3C, 4C, 5A, 5B	⊗	⊗	⊗	×	Blocked items—(3C, 4C) consistent-(5A) consistent-(5B) $E \propto 1/r$.
5A-(3C, 5B, 4C)-7C-24B				⊗	Blocked items—see text. (7C, 24B)—Newton’s 3rd law misconceptions.
4D, 5E	⊗	⊗	×	⊗	Blocked items—consistent.
14E-7C-24B	⊗				Newton’s 3rd law misconceptions.
8A, 9A	⊗				Charge on axis produces zero field.
8D, 9D	⊗	×	⊗	⊗	Interaction between charges modifies superposition.
9C-8D-9D		⊗			Interaction between charges modifies superposition.
10B, 21A		⊗		⊗	(10B, 21A) Velocity proportional to applied force. (21A) Electric-magnetic field indiscriminated.
10E, 11B	⊗	⊗	⊗	⊗	Blocked item—consistent.
13A, 14A	×		⊗		Conductor does not shield electric field.
13A—14A—13B	⊗				Conductor does not shield electric field—(13B) Like charges attract.
17C, 18C	⊗				Equipotential spacing proportional to electric field (18C); Electric field proportional to work (17C).
18E, 20C	⊗	⊗	×		Electric field-potential indiscriminated.
19B, 20B	×	×	⊗	×	Electric field points to higher potential. Electric field points to higher potential. (20A) Electric field-potential indiscriminated. (20A) Electric field proportional to equipotential spacing.
20A-19B-20B	⊗	⊗		⊗	
23C, 26B	⊗				Electric-magnetic field indiscriminated.
23D, 26C	⊗		⊗		Left hand rule.
Mixed correct and incorrect communities					
5D-(4B*, 5C*, 3B*)-7C-4D-5E			⊗		Blocked items. See text.
(6E*, 8B*, 9B*, 10C*)-21B				⊗	(10C*, 21B) Electric-magnetic field indiscriminated.
13E*, 14B			⊗		Shielding works symmetrically.
17B-18D*-20D*			⊗		Work proportional to electric field.
17E*-18E-20C-19A*			⊗		(18E, 20C) Electric field-potential indiscriminated.
19A*, 20C			×	⊗	(20C) Electric field-potential indiscriminated. (19A*, 20C) Field points to lower potential.
Completely correct communities					
(3B*, 4B*, 5C*, 7B*)-13E*				⊗	Coulomb’s law.
8B*, 9B*			⊗	×	Coulomb’s law and superposition.
15A*-17E*-16E*-(25D*, 21E*, 24C*, 27E*)				⊗	See text.
18D*, 20D*			×	⊗	Relation of equipotential spacing to field.
21E*, 27E*			⊗	×	Zero velocity implies zero magnetic force.
23A*, 26A*			⊗	⊗	Magnetic field of wires.

found that items within item blocks often were found in the same community in the FCI and FMCE. In Table I, communities largely composed of blocked items are labeled blocked items. When the responses to later items would have been correct if the response to the earlier item was correct, the community is labeled “consistent.”

Many correct, mixed, and incorrect communities contain combinations of responses to items 3, 4, and 5. The common stem of this set of items presents the student with two objects each with charge Q which exert a force F on each other. The charge of one of the objects is then increased to $4Q$. Item 3 asks about the new force on the object of charge Q . Item 4 asks about the new force on the object of charge $4Q$. Item 5 asks about how the force on the $4Q$ charge changes if the objects are moved 3 times farther apart. Many of the later responses in these communities are the correct response if an earlier response were correct as in communities {3A, 4A}, {3C, 4C, 5A}, and {4D, 5E}. The community {4D, 5E} is found independently and as part of a larger mixed community. The responses to item 3 in these communities all represent a failure to understand the relation of electric charge to electric force. The correct answers to items 3, 4, 5 were found both in a mixed and a completely correct community. In the mixed community, 5D is also associated with the correct responses (3B*, 4B*, 5C*); students selecting this response are applying the $E \propto 1/r$ misconception. Response 5B in the community {3C, 4C, 5A, 5B} also applies the $E \propto 1/r$ misconception. Response 5D in community {3D, 5D} may also apply this misconception except the student mistakenly selected $F/4$ instead of $4F$ for item 3 (this is the most straightforward way to reconcile these two inconsistent responses). Communities mixing items 3, 4, and 5 with other items were also identified and are discussed later.

Blocked items 10 and 11 are only found together in one incorrect community, {10E, 11B}; response 11B is the correct response if response 10E were correct. As noted in prior work, the practice of blocking items generates relations, correlations, between items that make the score of the items difficult to interpret [12–14]. A simple correct or incorrect scoring of items 4, 5, and 11 almost certainly understates a student’s understanding of the items; a modified scoring rubric that takes into account relations between the items is proposed in Sec. III J.

B. Mechanics communities

Multiple communities were identified where the responses represented incorrect reasoning about Newtonian mechanics. Students continue to apply some non-Newtonian reasoning identified with the FCI and FMCE in introductory mechanics courses in introductory electricity and magnetism courses. Three responses, 7C, 14E, and 24B, as well as inconsistent responses to items 3 and 4 show that Newton’s 3rd law is not well understood.

Hestenes and Jackson [11] identified two Newton’s 3rd law misconceptions in the FCI; greater mass implies greater force and most active agent produces greater force. While neither is completely appropriate for the CSEM, some responses to item 4 and responses 7C and 24B seem more aligned with the greater mass implies greater force misconception; response 7C applies a greater charge implies greater force reasoning, while response 24B applies greater current implies greater force reasoning. Response 14E, which involves an asymmetric application of the shielding by a conductor of the electric field, does not fit the larger implies more force model. All three responses indicate that the student does not apply Newton’s 3rd law in a variety of contexts. It also possible that students’ thinking is better modeled by the knowledge-in-pieces framework where the student is applying a p prim such as “increased effort begets greater results.” This is supported by the application of the same incorrect reasoning in multiple disparate contexts, but is not supported by the connection of Newton’s 3rd law linking all items. Further qualitative research is needed to resolve the correct model of student thinking on these items.

Response 7C was also associated with different combinations of the item 3, 4, 5 block in completely incorrect community {5A-(3C, 5B, 4C)-7C-24B} and mixed community {5D-(4B*, 5C*, 3B*)-7C-4D-5E}. The inclusion of 7C in these communities is unusual; item 4 can be answered either using Newton’s 3rd law or Coulomb’s law. Response combinations {3B*, 4B*} and {3C, 4C} represent a correct application of Newton’s 3rd law, though in the latter Coulomb’s law is not applied correctly. The mixing of incorrect applications of Newton’s 3rd law, 7C and 24B, with its correct application provides evidence that students are not applying a robust alternate theory and that Newton’s 3rd law reasoning in electricity and magnetism may be better modeled using p prim.

The responses in the community {10B, 21A} apply the force proportional to velocity misconception identified in Hestenes and Jackson’s taxonomy [11].

C. Isomorphic items

Study 2 identified 3 groups of isomorphic items {6, 8}, {16, 17}, and {21, 27}. Isomorphic items require very similar reasoning for their solution. In both study 3 and 4, isomorphic items often formed both incorrect and correct communities. The isomorphic items were less important in forming communities in the CSEM. All three sets of isomorphic items were identified together in a completely correct community or as correct items in a mixed community. Only items 21 and 27 were identified independently as a correct community; there was no corresponding incorrect community. These items ask for the magnetic force on a stationary charge; students are consistently reasoning correctly, but do not apply consistent incorrect reasoning. For these items, the most likely misconception or p prim

that zero velocity implies zero force also yields the correct answer.

While items 6E* and 8B* were identified in a mixed correct community, item 8 was more consistently associated with item 9 in incorrect communities. These two items may be described as nearly isomorphic; item 8 involves the change in electric force on a point charge as a third charge is added to the system; item 9 also uses two point charges and adds a third asking about the electric field. In study 2, the solution of the items differ only by the application of the relation between the force and field ($\vec{F} = q\vec{E}$). The point charge experiencing the force is a positive charge, and therefore force and field are parallel. Responses 8D, 9C, and 9D were found in incorrect communities in all samples; all explicitly test the misconception that the addition of an additional charged object somehow modified the total field beyond simply adding the field of the new object. It seems likely this misconception is responsible for the formation of the communities that include items 8 and 9.

Items 23 and 26 differ only by the principle of superposition. Item 23 asks about the addition of the fields of two wires; item 26 asks for the direction of the field of a single wire. Study 2 found item 23 discriminates very weakly on the principle of superposition; as such, these items may be effectively isomorphic explaining their identification as a correct community {23A*, 26A*} and as an incorrect community {23D, 26C} where the students incorrectly apply the right hand rule.

Items 16 and 17 are only found in one correct community connecting disparate items. Both are coded as applying the definition of electric potential in study 2. While this is true, the items are fairly different with item 16 applying the principle that only differences in electric potential are physically important while item 17 asks to rank the work needed to move through a field where the equipotentials are given. Study 2 did not include the principle “only differences in electric potential are physically relevant” in its multidimensional IRT models; it seems likely, given the failure to find items 16 and 17 in the same communities, that this was an oversight and they may not actually be isomorphic.

D. Electric-magnetic field indiscriminated

Study 1 reported that students often conflated electric fields and magnetic fields in items 23 and 26. The community {23C, 26B} is consistent with their analysis. The wires in these items seemed to be viewed as charges with magnetic field pointing either away from or toward the wire in the incorrect community. There could be a reverse conflation of electric fields and magnetic fields in the community {10E, 11B} where students seem to apply the Lorentz force law to electric fields. This connection is less clear; response 10E states that a charged particle released in an electric field remains at rest. There could be many reasons for the selection of this response. Mixed

community {10C*, 21B} also contains responses that were answered symmetrically for the electric and magnetic fields, where a uniform field produces a constant acceleration on a charge released in the field.

E. Electrostatic communities

Items 13 and 14 appear in two completely incorrect communities, {13A, 14A} and {13A-14A-13B}, and one mixed community {13E*, 14B}; these communities demonstrate different incorrect ideas about conductors and shielding. The two completely incorrect communities show the student does not understand that conductors shield the electric field. The mixed community applies the principle of shielding symmetrically reasoning that conductors not only shield their interior from external electric fields, but their exterior is also shielded from internal electric fields.

The completely correct community {(3B*, 4B*, 5C*, 7B*)-13E*} is curious. Items 3 to 7 all apply Coulomb's force law; however, the connection of this group with item 13 is unclear. Response 13E* correctly applies the principle of the shielding of a conductor's interior from electric fields and forces.

F. Electric potential communities

Items 16 to 20 all require the definition of electric potential. Item 16 was discussed as part of the isomorphic group formed of items 16 and 17 and is only identified in one community which mixed electric potential and magnetostatic items. The remaining electric potential items are found in a large number of communities. The composition of these communities can shed light on how these items are functioning. The completely correct community {18D*, 20D*} includes responses that correctly represent the relation of equipotential spacing and field strength, 18D* and 20D*, and the relation of equipotential magnitude and electric field direction, 20D*. Mixed community {17B-18D*-20D*} connects 17B to 18D* which may indicate the student believes work is proportional to electric field independent of distance. The mixed community {19A*, 20C} is identified independently and as part of the larger community {17E*-18E-20C-19A*}. Both responses 19A* and 20C correctly capture the concept that electric field points toward lower electric potential; response 20C applies the electric field-potential indiscriminated naive conception. Responses consistent with the idea that the electric field is proportional to the electric potential may indicate a student applied this naive conception. This implies a student being scored as incorrect on 20C is demonstrating some of the knowledge the correct students are demonstrating. This pair is connected with response 18E in mixed community {17E*-18E-20C-19A*}; 18E also represents the electric field-potential indiscriminated misconception. The combination {18E-20C} is also identified as an independent community. Response 17E* is the correct response that work is

proportional to the difference in potential, but the connection to 18E suggests the student does not understand the relation of electric field to potential and may be reasoning that work is proportional to electric field. The connection of the correct response to the incorrect response suggests the item 17 may not be functioning as intended.

Many completely incorrect communities that are not part of mixed communities were also identified. In community {17C, 18C}, a student selecting 18C applies the reasoning that electric field magnitude increases with equipotential spacing; this suggests that the student's selection 17C may result from reasoning that work is proportional to electric field. All items in the community {20A-19B-20B} suggest the student believes electric field points to higher potential. Item 20A could also represent the idea that electric field magnitude increases with equipotential spacing or the electric field-potential indiscriminated naive conception. Applying the "increased effort begets greater results" p prim could also lead a student to reason that the electric field magnitude increases with equipotential spacing.

G. Magnetostatic communities

Incorrect communities {23D, 26A} and correct communities {21E*, 27E*}, {23A*, 26A*} were discussed earlier as examples of isomorphic or nearly isomorphic item groups; response 24B was discussed as applying Newton's 3rd law misconceptions. Response 21B was identified as applying the electric-magnetic field indiscriminated misconception. The only other community formed of magnetostatic items combined these items with electric potential items and is discussed in the following section.

H. Other communities

The community {15A*-17E*-16E*-(25D*, 21E*, 24C*, 27E*)} mixes electric potential and magnetostatic items. The physical reasoning linking these items is unclear. One possible explanation for this community is

that many of the responses are of a type preferentially avoided by students [52] for nonphysical reasons. Response 16E is a "none of the above" response, responses 17E and 24C report all the quantities are equal, response 21E and 27E are "zero" responses, and response 25D have the quantities in ascending order.

I. Misconception scores

To help instructors to extract more information from the application of conceptual instruments, both study 3 and Wells *et al.* [7] calculated misconception scores indicating how strongly students held certain misconceptions. Misconception scores are presented in Table II. A misconception score is a percentage representing the number of responses a student selected that are identified with a given misconception out of the total number of CSEM responses identified with that misconception. For example, for the conductor and insulator misconceptions (1C, 2B), an individual student can select 0, 1, or 2 of these responses and have a misconception score of 0%, 50%, or 100%. The 10.0% misconception score of sample 1 for this misconception represents the average of the student misconception scores for that sample. This work refines the calculation of misconception scores by identifying combinations of items which taken together represent a misconception; these combinations are counted as a single instance of a misconception and are represented by two items in parenthesis in Table II. Multiple combinations of responses to items 3 and 4 represent a failure to apply Newton's 3rd law correctly; one misconception is counted for each time the response to item 3 is different from the response to item 4.

Communities with a very low misconception score such as {8A, 9A} and {23D, 26C} contain responses chosen by very few students. The correlation coefficient, which is used to define the relation between responses, is fairly insensitive to sample size. The identification of these communities selected by few students implies those

TABLE II. Misconception scores. Items not in parentheses represent independent misconceptions. Items in parentheses represent a misconception only if both items are selected and are counted as one response.

Misconception	Responses	Misconception score	
		Sample 1	Sample 2
Conductor and insulator misconceptions	1C, 2B	10.0%	18.1%
Electric field proportional to $1/r$	(3A, 4D), (4B, 5D), (4C, 5D)	12.8%	17.2%
Newton's 3rd law misconceptions	(3 $\neq$ 4), 7C, 14E, 24B	17.6%	26.5%
Charge on axis produces zero field	8A, 9A	2.5%	4.4%
Interaction between charges modifies superposition	8D, 9C, 9D	17.3%	31.1%
Velocity proportional to applied force	10B, 21A	12.9%	24.5%
Shielding misconceptions	13A, 13B, 14A	27.8%	56.2%
Electric field-potential indiscriminated	18E, 20C	32.3%	31.4%
Electric-magnetic field indiscriminated	(10B, 21A) (10C*, 21B), 23C, 26B	2.2%	13.3%
Electric field points to higher potential	19B, 20A, 20B	13.0%	29.5%
Left hand rule	23D, 26C	4.6%	8.7%

students who select one response in the community also selected the other response.

Sample 2 had consistently higher misconception scores than sample 1 for all categories excluding electric field-potential undiscriminated, which was commensurate for both samples. This was expected given the average CSEM score for each sample. The high misconception score of 56.2% for sample 2's shielding misconceptions category along with the lack of communities composed of items 13A, 13B, and 14A in sample 2's networks could be explained by a general misunderstanding of shielding by conductors. The high misconception score for both samples' electric field-potential undiscriminated naive conception as well as the community appearing in both samples' networks indicates that the conflation of electric fields and potentials could be a broadly held naive conception among many students postinstruction.

J. Alternate scoring rubric

The consistent answering patterns on blocked items and the existence of isomorphic groups of items suggest an alternate scoring rubric for the CSEM might be appropriate. Items 3, 4, and 5 are blocked items all measuring the understanding of Coulomb's force law. We propose item 4 be counted as correct if it is consistent with the response to item 3 or is itself correct. Likewise, item 5 should be counted as correct if it is answered consistently with item 4 or is itself correct. However, repeated instances of the response E, "other," for items 3, 4, and 5 should not be considered consistent for the grading of items 3, 4, and 5. Items 10 and 11 are blocked items involving the motion of a charged particle in a uniform electric field. Response 11B should be counted as correct if response 10E was selected; both assert the particle remains at rest. Items 21 and 27 both test whether the student understands that the magnetic force on a stationary charge is zero; we suggest grading these two items as a block where one point is received if both are correct. This would also serve to lower the weighting of this relatively specific piece of knowledge on the overall score.

Item 4 could either be answered using the same reasoning as item 3 using Coulomb's law or by applying Newton's 3rd law to the result of item 3. The majority of communities identified involving item 4 do not consistently contain the other Newton's 3rd law items. It seems reasonable to take consistent reasoning on this pair as evidence of correct reasoning based on a prior incorrect answer.

An alternate rubric for the BEMA is available at PhysPort [10]. CSEM items 3 to 5 are very similar to items 1 to 3 on the BEMA; the alternate scoring of item 5 suggested is consistent with the suggestions for the BEMA. The modified rubric to the BEMA also suggests grading one pair of isomorphic items as a block where the student receives one point if both are correct.

With the modified scoring producing a score out of 31, the Sample 1 CSEM percentage score became 62.2 ± 0.3 and the Sample 2 CSEM percentage score became 46.4 ± 0.3 . The modified scoring resulted in a 0.4 percentage point increase in average CSEM score for sample 1 and a 0.9 percentage point increase in average score for sample 2. The small change in overall score is the result of a decrease in score resulting from treating 21 and 27 as a block partially cancelling the increase in score resulting from giving credit for consistent answering of 3, 4, 5, 10, and 11.

IV. DISCUSSION

A. Research questions

This work explored three research questions which will be addressed in the order proposed.

RQ1: What community structure is identified by network analysis of the CSEM? How are the communities associated with previously identified features of the instrument? The community structure identified by MMA and MMA-P was discussed in detail in the previous section. Communities of incorrect responses or mixed correct and incorrect responses were identified in electrostatics, electric potential, and magnetostatics but not in magnetic induction. Most communities connected items within the individual topics; however, communities of items incorrectly applying Newtonian mechanics and communities applying the same reasoning to electric and magnetic fields included items from multiple topics. This indicates Newtonian misconceptions continued to be applied to multiple contexts involving both electricity and magnetism. The isomorphic items identified in study 2 were much less important to the community structure than in the FCI or FMCE in studies 3 and 4. This indicates the consistent incorrect reasoning patterns were more independent from the correct reasoning in electricity and magnetism than they were in mechanics.

Blocked items were included in many communities offering additional evidence that the practice of blocking items makes an instrument difficult to interpret statistically. Unlike prior studies, analysis of patterns of blocked responses were useful in further understanding the answering patterns of students. This suggests a simple all or nothing scoring of each CSEM item may fail to correctly capture the student's understanding of electricity and magnetism. An alternate scoring scheme that grades items contingently based on the responses to earlier items might produce a more accurate reflection of student understanding. One such scheme was proposed in Sec. III J. While some individual students scores were modified, the overall CSEM average changed little.

Many of the communities identified represented consistent application of incorrect reasoning identified in the introduction of the instrument in study 1. Network analysis showed that students were applying similar incorrect

reasoning across multiple contexts. Some communities represented failure to broadly understand general concepts: shielding misconceptions and conductor-insulator misconceptions. Some represented naive conceptions where two disparate quantities were assumed to behave in the same manner: electric-magnetic field undiscriminated, electric field-potential undiscriminated, or electric field proportional to work. Some seemed to represent simple mistakes: incorrectly applying the right-hand rule or believing the electric field points toward higher potential. There were also a few misconceptions that assumed additional properties of the electric field not consistent with physics: interaction between charges modifies superposition and a charge on the axis produces zero field. It seems likely that all these general types of consistent incorrect reasoning are also present in mechanics but are not detected by the FCI which was developed explicitly to detect common misconceptions. In summary, the substantial variety of incorrect reasoning measured by the CSEM may indicate it is a superior instrument to the FCI for exploring the structure of incorrect physical understanding.

Most completely correct communities involved items requiring either the same reasoning or very similar reasoning. Many of the mixed communities involved chains of items where members of the community are only connected in pairs. Many of the mixed communities involved electric potential. Most of these communities were only identified in sample 1. This may indicate that electric potential is not being effectively covered for the students in sample 1; this is supported by a misconception score for the electric field-potential undiscriminated naive conception in Sample 1 that is much more similar to the weaker performing sample 2 than other misconception scores. The mixing of correct and incorrect responses for electric potential in sample 1 may indicate the correct responses are being selected without correct understanding and that the scores on these items may overstate student knowledge. Some mixed communities were also formed when an incorrect response required multiple reasoning steps; the student may reason correctly on one step and incorrectly on another step ultimately selecting the incorrect response. This can be seen in the community {19A*, 20C} where the student correctly reasons that electric field points to lower potential but does not understand the relation between equipotential spacing and field strength.

RQ2: Does the community structure of the CSEM have communities related to Newtonian mechanics? If so, how do these communities compare to the communities identified in the FCI or the FMCE? Two general groups of mechanics misconceptions were identified in the CSEM by module analysis. The first, responses {10B, 21A}, applied the velocity proportional to applied force misconception and was identified by both MMA and MMA-P, but only in the lower performing sample 2. Communities testing this misconception were also identified in the FMCE by Wells

et al. [7]. These responses had moderate misconception scores of 12.9% for sample 1 and 24.5% for sample 2. As such, while this misconception does persist after instruction in mechanics, it is not one of the most applied misconceptions in the electricity and magnetism class.

Four items are potentially related to an incorrect application of Newton's 3rd law: 7C, 14E, 24B, and inconsistent responses to items 3 and 4. These items were identified in communities in a number of combinations. All communities containing 24B also contain 7C. Two of the three communities containing 7C also contain 24B. The only community containing 14E also contains 7C and 24B. Response 7C is also found in communities containing incorrect responses to item 4, either response 4C or 4D. Responses 7C and 24B are consistent with a greater charge/current implies greater force misconception which is analogous to the "greater mass implies greater force" misconception in mechanics. Response 14E only states the forces are different; the relative charges of the two objects involved are not given and, therefore, it is not possible to determine if a modification of greater mass implies greater force is being applied. The inconsistent association of 7C to responses to item 4 and the failure to associate either responses 14E or 24B with these items seems to indicate inconsistent responses to items 3 and 4 are selected for reasons other than applying Newton's 3rd law incorrectly. The communities formed by all four items are very different than the communities of Newton's 3rd law items identified in study 3 and Wells *et al.* [7] which were tightly interconnected. The Newton's 3rd law responses in the current study were sparsely connected. This seems to indicate that students' application of incorrect reasoning about Newton's 3rd law is far less consistent when applied to electricity and magnetism or that the reasoning is not well modeled as a misconception but rather should be thought of as the application of a *p* prim.

RQ3: How do the communities identified by the two versions of Module Analysis, MMA and MMA-P, compare? How do the communities identified at different institutions compare? A total of 20 completely incorrect communities were identified by MMA or MMA-P in the two samples; only 6 were identified by both techniques in both samples. Four of these were responses to blocked items. MMA identified 19 communities in the two samples, 8 were identified in both samples. Sixteen communities were identified in the higher performing sample 1, while only 11 were identified in the lower performing sample 2. All MMA-P communities identified in sample 1 were also identified as MMA communities. The 8 communities identified by MMA-P in sample 1 were substantially fewer than the 16 identified by MMA suggesting many of the communities identified by MMA in sample 1 resulted from correlations through total test score. This, and the facility to include correct answers, suggests MMA-P is a superior technique for exploring consistent student answering patterns than MMA.

For sample 2, only one fewer community was identified by MMA-P than MMA. All MMA-P communities in sample 2 were also identified by MMA except {5A-(3C,5B,4C)-7C-24B}. The additional community is curious because it mixed incorrect application of Newton's 3rd law in responses 7C and 24B with a correct application in 4C based on the incorrect response in 3C. As such, the MMA-P algorithm produced a simplified community structure and removed many communities which were correlated through total CSEM score.

There was very little similarity between the mixed communities in the two samples; only one community was identified in both samples. Many more mixed communities were identified in sample 1. Both mixed communities for sample 2 involved confounding different concepts through the electric-magnetic field undiscriminated and the electric field-potential undiscriminated naive conceptions. Application of naive conceptions is consistent with the overall low performance of the sample and shows these students continue to have a fundamental misunderstanding of the central concepts of electricity and magnetism postinstruction.

Four of the six completely correct communities were consistent between samples. One of the inconsistent communities {15A*-17E*-16E*-(25D*,21E*,24C*,27E*)} may be related to issues of test logic such as avoiding none of the above responses with little relation to physical reasoning. The other inconsistent community is curious {(3B*,4B*,5C*,7B*)-13E*}; four of the responses are related to Coulomb's force law while response 13E* involves understanding that a conductor shields its interior from the electric field.

B. Other observations

This analysis identified many disparate reasoning patterns. Many incorrect communities resulted from the naive application of reasoning associated with one concept to a different concept: electric-magnetic field undiscriminated and electric field-potential undiscriminated. A number of communities seem to result from simple mistakes which seem unlikely to represent strongly held beliefs: electric field points to higher potential and the left hand rule. Other communities do not seem to represent a single consistent fragment of incorrect reasoning, but more general topics that are not well understood: conductor and insulator misconceptions and electric shielding by conductors. Many communities could be explained by an alternate framework of student knowledge. The increased effort begets greater results p prim might explain the electric field-potential undiscriminated, the electric field points to higher potential, the electric field proportional to work, and the Newton's 3rd law misconceptions.

It seems unlikely that any of the incorrect reasoning about electricity and magnetism resulted from the students' lived experiences. It also seems likely that similar

differences in strength and kind of incorrect reasoning are found in mechanics. It is important to understand these differences because effective and efficient methods of addressing incorrect thinking may differ depending on the origin and degree the student has internalized the misconception. For example, a misconception rooted in a robust naive theory applicable and productive across many contexts may require substantial instructional resources to overcome; however, a simple mistake such as using your left hand to do a cross product might be corrected by simply pointing out the mistake.

The variety of coherent incorrect responses identified in this study suggests that an instrument like the FCI developed to measure strongly held misconceptions may miss a variety of other incorrect answering patterns. Evidence supporting this possibility can be found in the MMA analysis of the FMCE by Wells *et al.* [7]. Both the CSEM and FCI use five responses per item where the responses were developed from interviews and student responses to open-response applications of the instrument. The FMCE employs a different strategy offering the student up to nine responses per item where the items generally exhaust all possible responses. The present study and the MMA analysis of the FCI in study 3 both discarded nodes selected by fewer than 30 students producing relatively compact disconnected communities with a correlation threshold of 0.15 or 0.20. This was not the case for the FMCE. When Wells *et al.* removed only nodes selected by fewer than 30 students, an exceptionally complex community structure was produced at the $r > 0.20$ correlation threshold; to reproduce the compact community structure of the FCI, only nodes selected by 20% of the students were retained. When this threshold was relaxed to 10% of the students, a substantially more complex community structure was exposed where the communities generally could be identified with a pattern of incorrect reasoning. Often this incorrect reasoning was not represented as a misconception in Hestenes and Jackson's taxonomy [11]. The work on the FMCE and the present work on the CSEM suggest there is a much richer ecology of incorrect conceptions than those measured by the incorrect reasoning on an instrument such as the FCI which focuses on the most strongly held misconceptions. The relations identified between items within the CSEM suggested an alternate scoring rubric was needed to correctly represent student understanding; one such rubric was proposed.

V. IMPLICATIONS

Network analysis was successful in identifying consistently selected incorrect and correct answers in the CSEM. This provides support for the need for a taxonomy of the common incorrect answering patterns of electricity and magnetism similar to the taxonomy of mechanics misconceptions provided by Hestenes and Jackson [11]. The variety of incorrect communities identified in the CSEM was

commensurate with the variety of incorrect communities measured by the FCI in study 2 and larger than the number of communities found in the FMCE by Wells *et al.* [7]. This suggests the existence of a rich set of patterns of incorrect reasoning about electromagnetism; a systematic identification of the full set of these patterns would benefit instructors. The calculation of misconception scores similar to those in Table II should allow instructors to determine what kinds of incorrect reasoning most need to be addressed.

VI. FUTURE WORK

This work provided qualitative comparisons of the CSEM networks from two institutions. In the future, quantitative measures of network similarity may yield additional insights. The communities in Fig. 1 demonstrate a diverse pattern of connections; some are linear while some are fully connected polygons or mixes of the two. Additional research is needed to determine if the morphology of the community further informs understanding of student reasoning.

VII. CONCLUSION

The CSEM was constructed to assess students' conceptual understanding of electricity and magnetism. This study explored the use of MMA and MMA-P as productive ways to identify communities of correlated responses to individual items within the CSEM. In general, MMA-P produced a richer set of communities and eliminated communities formed of items correlated through total instrument score. As such, while both techniques yield useful results, MMA-P seems the superior technique for exploring conceptual understanding. Overall, a number of communities were identified for the two samples; however, the explanations for the reasoning represented by these communities varied. A substantial number of the identified

incorrect communities consisted of blocked items providing continued support that the practice of blocking items can produce correlations that are not related to physical reasoning. The consistent identification of blocking as generating psychometric problems for the primary conceptual instruments used in PER may suggest the need for a new generation of conceptual inventories. Multiple communities were formed of responses where the response to later items would be correct if the response to an earlier item was correct. This suggests that the scoring rubric to the CSEM should be modified to include relations between responses. A modified scoring rubric was proposed, but changed overall CSEM post-test averages little. Most communities of completely incorrect responses and mixed correct and incorrect responses consisted of items with the same subtopic, either electrostatics, electric potential, or magnetostatics. Some communities connected items in multiple subtopics including misconceptions about mechanics and a failure to differentiate the electric and the magnetic field. The results suggest the existence of a rich collection of misconceptions and naive conceptions about electricity and magnetism which are consistently applied by students after instruction in introductory physics. This collection was much more diverse than those identified by the FCI, which may indicate that the FCI does not fully characterize the scope of coherent incorrect reasoning.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation as part of the evaluation of improved learning for the Physics Teacher Education Coalition, PHY-0108787. Data collection for this work was supported by National Science Foundation Grants No. EPS-1003907 and No. ECR-1561517.

-
- [1] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
 - [2] 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).
 - [3] D. P. Maloney, T. L. O'Kuma, C. Hieggelke, and A. Van Huevelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
 - [4] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief Electricity and Magnetism Assessment, *Phys. Rev. Phys. Educ. Res.* **2**, 010105 (2006).
 - [5] E. Brewster, J. Bruun, and I. G. Bearden, Using module analysis for multiple choice responses: A new method applied to Force Concept Inventory data, *Phys. Rev. Phys. Educ. Res.* **12**, 020131 (2016).
 - [6] J. Wells, R. Henderson, J. Stewart, G. Stewart, J. Yang, and A. Traxler, Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis, *Phys. Rev. Phys. Educ. Res.* **15**, 020122 (2019).
 - [7] J. Wells, R. Henderson, A. Traxler, P. Miller, and J. Stewart, Exploring the structure of misconceptions in the Force and Motion Conceptual Evaluation with modified module analysis, *Phys. Rev. Phys. Educ. Res.* **16**, 010121 (2020).

- [8] J. Yang, J. Wells, R. Henderson, E. Christman, G. Stewart, and J. Stewart, Extending modified module analysis to include correct responses: Analysis of the Force Concept Inventory, *Phys. Rev. Phys. Educ. Res.* **16**, 010124 (2020).
- [9] C. Hieggelke and T. O’Kuma, The impact of physics education research on the teaching of scientists and engineers at two-year colleges, *AIP Conf. Proc.* **399**, 267 (1997).
- [10] Physport, <https://www.physport.org>. Accessed 8/30/2020.
- [11] Table II for the Force Concept Inventory (revised from 081695r), http://modeling.asu.edu/R&E/FCI-RevisedTable-II_2010.pdf. Accessed 3/17/2019.
- [12] J. Stewart, C. Zabriskie, S. DeVore, and G. Stewart, Multidimensional item response theory and the Force Concept Inventory, *Phys. Rev. Phys. Educ. Res.* **14**, 010137 (2018).
- [13] J. Yang, C. Zabriskie, and J. Stewart, Multidimensional item response theory and the Force and Motion Conceptual Evaluation, *Phys. Rev. Phys. Educ. Res.* **15**, 020141 (2019).
- [14] C. Zabriskie and J. Stewart, Multidimensional item response theory and the Conceptual Survey of Electricity and Magnetism, *Phys. Rev. Phys. Educ. Res.* **15**, 020107 (2019).
- [15] K. A. Zweig, *Network Analysis Literacy: A Practical Approach to the Analysis of Networks*, Lecture Notes in Social Networks (Springer Vienna, Vienna, Austria, 2016).
- [16] D. P. Maloney, Charged poles?, *Phys. Educ.* **20**, 310 (1985).
- [17] C. Guruswamy, M. D. Somers, and R. G. Hussey, Students’ understanding of the transfer of charge between conductors, *Phys. Educ.* **32**, 91 (1997).
- [18] S. Törnkvist, K. A. Pettersson, and G. Tranströmer, Confusion by representation: On student’s comprehension of the electric field concept, *Am. J. Phys.* **61**, 335 (1993).
- [19] I. Galili, Mechanics background influences students’ conceptions in electromagnetism, *Int. J. Sci. Educ.* **17**, 371 (1995).
- [20] L. Viennot and S. Rainson, Students’ reasoning about the superposition of electric fields, *Int. J. Sci. Educ.* **14**, 475 (1992).
- [21] S. Rainson, G. Tranströmer, and L. Viennot, Students’ understanding of superposition of electric fields, *Am. J. Phys.* **62**, 1026 (1994).
- [22] M. Planinic, Assessment of difficulties of some conceptual areas from electricity and magnetism using the Conceptual Survey of Electricity and Magnetism, *Am. J. Phys.* **74**, 1143 (2006).
- [23] K. Kreutzer and A. Boudreaux, Preliminary investigation of instructor effects on gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010120 (2012).
- [24] P. B. Kohl and H. V. Kuo, Introductory physics gender gaps: Pre-and post-studio transition, *AIP Conf. Proc.* **1179**, 173 (2009).
- [25] J. M. Wilson, The CUPLE physics studio, *Phys. Teach.* **32**, 518 (1994).
- [26] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, Exploring the gender gap in the Conceptual Survey of Electricity and Magnetism, *Phys. Rev. Phys. Educ. Res.* **13**, 020114 (2017).
- [27] A. Madsen, S. B. McKagan, and E. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, *Phys. Rev. Phys. Educ. Res.* **9**, 020121 (2013).
- [28] J. Leppävirta, The effect of naïve ideas on students? Reasoning about electricity and magnetism, *Res. Sci. Educ.* **42**, 753 (2012).
- [29] D. E. Meltzer, Analysis of shifts in students’ Reasoning regarding electric field and potential concepts, *AIP Conf. Proc.* **883**, 177 (2007).
- [30] N. I. Karim, A. Maries, and C. Singh, Exploring one aspect of pedagogical content knowledge of teaching assistants using the Conceptual Survey of Electricity and Magnetism, *Phys. Rev. Phys. Educ. Res.* **14**, 010117 (2018).
- [31] J. Clement, Students’ preconceptions in introductory mechanics, *Am. J. Phys.* **50**, 66 (1982).
- [32] J. Clement, D. E. Brown, and A. Zietsman, Not all preconceptions are misconceptions: Finding “anchoring conceptions” for grounding instruction on students’ intuitions, *Int. J. Sci. Educ.* **11**, 554 (1989).
- [33] J. Clement, Using bridging analogies and anchoring intuitions to deal with students’ preconceptions in physics, *J. Res. Sci. Teach.* **30**, 1241 (1993).
- [34] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985).
- [35] I. A. Halloun and D. Hestenes, Common sense concepts about motion, *Am. J. Phys.* **53**, 1056 (1985).
- [36] M. T. H. Chi and J. D. Slotta, The ontological coherence of intuitive physics, *Cognit. Instr.* **10**, 249 (1993).
- [37] M. T. H. Chi, J. D. Slotta, and N. De Leeuw, From things to processes: A theory of conceptual change for learning science concepts, *Learn. Instr.* **4**, 27 (1994).
- [38] J. D. Slotta, M. T. H. Chi, and E. Joram, Assessing students’ misclassifications of physics concepts: An ontological basis for conceptual change, *Cognit. Instr.* **13**, 373 (1995).
- [39] A. A. diSessa, Toward an epistemology of physics, *Cognit. Instr.* **10**, 105 (1993).
- [40] A. A. diSessa and B. L. Sherin, What changes in conceptual change?, *Int. J. Sci. Educ.* **20**, 1155 (1998).
- [41] J. Minstrell, Facets of students’ knowledge and relevant instruction, in *Research in Physics Learning: Theoretical Issues and Empirical Studies*, edited by R. Duit, F. Goldberg, and H. Niedderer (IPN, Kiel, Germany, 1992), pp. 110–128.
- [42] D. Hammer, Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions, *J. Learn. Sci.* **5**, 97 (1996).
- [43] D. Hammer, More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research, *Am. J. Phys.* **64**, 1316 (1996).
- [44] D. Hammer, Student resources for learning introductory physics, *Am. J. Phys.* **68**, S52 (2000).
- [45] R. E. Scherr, Modeling student thinking: An example from special relativity, *Am. J. Phys.* **75**, 272 (2007).
- [46] A. A. diSessa, A friendly introduction to “knowledge in pieces”: Modeling types of knowledge and their roles in learning, in *Invited Lectures from the 13th International Congress on Mathematical Education*, edited by G. Kaiser,

- H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, and B. Xu (Springer International Publishing, Cham, 2018), pp. 65–84.
- [47] U. S. News & World Report: Education, <https://premium.usnews.com/best-colleges>. Accessed 4/30/2017.
- [48] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, NY, 1977).
- [49] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* **69**, 026113 (2004).
- [50] G. Csardi and T. Nepusz, The igraph software package for complex network research, *Inter Journal, Complex Systems* **1695**, 1 (2006).
- [51] A. Canty and B. D. Ripley, boot: Bootstrap R (S-Plus) Functions (2017), R package version 1.3–20.
- [52] S. DeVore, J. Stewart, and G. Stewart, Examining the effects of testwiseness in conceptual physics evaluations, *Phys. Rev. Phys. Educ. Res.* **12**, 020138 (2016).