Conceptual survey of electricity and magnetism: analysis of the items and recommendations for improvement

Pablo Barniol, Esmeralda Campos, and Genaro Zavala

Department of Physics, Tecnologico de Monterrey, E. Garza Sada 2501, Monterrey, Mexico, 64849

The Conceptual Survey of Electricity and Magnetism is an assessment tool widely used in the Physics Education Research community. Although it has been some time since its publication, there is no study to evaluate the test by analyzing its items in detail. This contribution aims to do that by performing three item-focused statistical tests: difficulty index, discrimination index, and point biserial coefficient. We use data from an American public university and a Mexican private university. We analyze two items (14 & 20) with the most critical index values in both populations. In both items we found that the main problematic aspect is that the required reasoning process to answer them is quite elaborate, and in item 14 we also found some misleading features of the wording and the figure. The results and discussion may be useful for researchers using the test as an assessment tool and for a possible future version of the test.

I. INTRODUCTION

Multiple-choice tests are often used in Physics Education Research because they allow for the evaluation of large populations [1]. Ding et al. [2] note that these instruments should be evaluated through statistic tests to assert their reliability and discriminatory power, and recommend performing three item-focused tests: difficulty index, discrimination index, and point biserial coefficient. As a result, some instruments have been improved based on these analyses [3].

In Electricity and Magnetism, the Conceptual Survey of Electricity and Magnetism (CSEM) [4] is one of the most widely used multiple-choice tests for conceptual evaluation. Although it has been some time since its publication, there is no study in which the instrument is evaluated by performing the three item-focused tests. This article aims to do that, with the specific objective of identifying the items with most critical index values to analyze them in detail, reveal their possible source of problems and then present recommendations in their use in the test.

II. PREVIOUS RESEARCH

In the study that introduces the CSEM, Maloney et al. [4] did not present a detailed analysis of the item difficulty and discriminatory indexes, they did not present the analysis of the item point-biserial coefficients either. In their item difficulty analysis, the researchers presented a histogram with all these indexes and detected only two general trends without elaborating: (1) "the items range in difficulty from about 0.10 to a little over 0.8, which is a reasonable range", and (2) "there are only about seven items with difficulties of 0.6 or larger, which is probably fewer than would be ideal." In their item discriminatory analysis, they also only detected two general trends without elaborating: (i) "students in a calculus-based physics classes had discrimination values ranging from 0.1 to 0.55 which are not as high as one would hope," and (ii) "all but four of the items had values greater than 0.2, which is the traditional lower limit for acceptability."

It is important to mention that Maloney et al.'s article of the CSEM was published long before Ding et al. [2], who described the three mentioned statistical tests, focusing on item analyses. Ding et al. also recommended a more conservative minimum value (0.3) for the difficulty and discriminatory indexes. Finally, Planinic [5] analyzed the instrument using the Rasch model analysis to a sample of 110 Croatian students, which revealed critical behaviors, especially of item 14. In the Discussion and Conclusion section we will compare these results with ours.

III. METHODS

A. Participants: American and Mexican students

To reduce the effect of the specific nature of the populations, we decided to use data of two different institutions: (1) a large public American research university (AMU), and (2) a large private Mexican research university (MXU). Students of both populations are from a calculus-based course of Electricity and Magnetism. In both universities students use well known textbooks ([6] and [7] respectively). Most students in both universities are engineering students. We used data from 219 students of AMU and 310 students of MXU. Following the recommendations by Henderson [8], we eliminated in both universities the students that left blank over 20% of the items. For MXU, we used a version of the CSEM in Spanish. Three physics instructors translated the test and any differences were discussed and reconciled.

B. Data analysis

To analyze the items, we followed the next procedure. First, following the recommendations suggested by Ding et al. [2], we calculated for both populations the three statistical tests that focus on individual assessment items: (1) the item difficulty index, (2) the item discriminatory index (using the 25%-25% method), and (3) the item point-biserial coefficient. The first test is the fraction of students who answer the item correctly, the second is a measure of discriminatory power of each item (between the students with high scores on the entire test and the students with low scores), and the third reflects the correlation between students' scores on an individual item and scores on the entire test. Then, we identified the most critical items, i.e., those that in both populations have the three index values below the minimum values recommended by Ding et al [2]. Subsequently, to reveal the problematic aspects of these critical items, we analyzed the proportion of students selecting each option, the required reasoning process based on physical principles to answer the item and its wording and figure.

IV. ANALYSIS OF THE CSEM ITEMS

A. Identification of the most critical items

While analyzing the item indexes, we found, in general, similar trends in the values of both populations. In the interest of conciseness, we do not include all the calculated indexes for both populations. The most critical items we identified in both populations are items 14 and 20: all three index values were found to be below the minimum recommended values, as shown in Table I. This means that these two items have three characteristics: (1) high difficulty for students, (2) absence of adequate discrimination between students with high scores on the entire test and the students with low scores, and (3) low correlation between items' scores and entire test scores. Regarding these items, note that in the study in which the CSEM is introduced, Maloney et al. [4] reported very low difficulty indexes for items 14 and 20 (0.16 and 0.21 respectively) for calculus-based students, however, the authors did not analyze these items.

TABLE I. Item indexes of the two items (14 & 20) that in both populations do not fulfill the minimum values recommended by Ding et al. [2] (AMU: American university; MXU: Mexican university).

Indexes & Desired values	Item 14		Item 20	
	AMU	MXU	AMU	MXU
Difficulty index [0.3,0.9]	0.17	0.21	0.25	0.19
Discriminatory index ($\geq$0.3)	0.07	0.18	0.20	0
Point-biserial coeff. ($\geq$0.2)	0.14	0.19	0.19	0.11

TABLE II. Percentage of students selecting each of the options in items 14 and 20 for both populations. The correct answer is boldface and N is for students that did not answer.

Item	Population	%					
		A	B	C	D	E	N
14	AMU	49	11	8	**17**	15	0
	MXU	43	10	10	**21**	16	0
20	AMU	12	27	29	**25**	6	1
	MXU	24	24	26	**19**	6	2

In the next two subsections we elaborate on the characteristics of these two items. Table II presents the percentage of students in both populations selecting each of the options for these items.

B. Analysis of item 14

Figure 1 shows the associated figure of item 14. The question is as follows:
14. The figure below shows an electric charge q located at the center of a hollow uncharged conducting metal sphere. Outside the sphere is a second charge Q. Both charges are positive. Choose the description below that describes the net electrical forces on each charge in this situation.
(a) Both charges experience the same net force directed away from each other.
(b) No net force is experienced by either charge.
(c) There is no force on Q but a net force on q.
(d) There is no force on q but a net force on Q.
(e) Both charges experience a net force but they are different from each other.
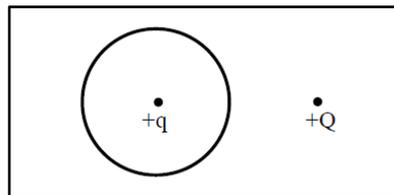
FIG 1. Figure of item 14 of the CSEM.

Item 14 is one of the most complicated items in the CSEM for both populations. As shown in Table II, the percentages of the correct answer for AMU and MXU students are very low (option D, 17% and 21% respectively). Analyzing Table II, we observe that the most frequent error for both populations is selecting option A. The percentages of this option are very high for both AMU and MXU students (49% and 43%). These percentages are very similar to those reported by Maloney et al. [4] and Planinic [5]. Maloney et al. mentioned that the selection of this error seems to be based on a misuse of Newton's third law, and Planinic stated that this error is based at least partly on Newton's third law. To analyze the possible source of problems of this item, we propose to follow the required reasoning process for this question (Table III).

TABLE III. Required reasoning process to answer item 14.

Step	Description of the reasoning
1	Realize that an "uncharged conducting metal sphere" is a "neutral conducting metal sphere" with equal numbers of positive and negative charges.
2	Realize that, since the question asks for "net electrical forces", all the charges of the system must be considered, i.e., the two point charges and the positive and negative charges of the sphere.
3	Realize that, based on the description of the object ("hollow sphere"), the sphere has inner and outer surfaces (although the figure does not clearly show them) and that the two point charges induce charge on both surfaces.
4	Note that for electrostatic equilibrium, i.e., zero electric field within the conductor, there has to be an induced uniformly distributed charge of $-q$ on the inner surface, and an induced non-uniformly distributed charge of $+q$ on the outer surface.
5	Conclude that the net force on the point charge $+q$ is zero because this charge is exactly at the center of the sphere, and that there is a non-zero net force on the point charge $+Q$ due to the non-uniformly distribution of charge on the outer surface of the sphere.

Analyzing this process, the item's wording and figure, and the selection of the most frequent error, we can identify the two main problematic aspects of this item that seem to explain the low indexes. The first aspect is that it requires, as shown in Table III, a very elaborate reasoning process. As we can observe, the process has five steps of reasoning and some of them, e.g., step 4, require the understanding of several electrostatic concepts. In another study [3], we noted that an item of a mechanical waves test with this characteristic, i.e., having several steps of reasoning, also showed a very low difficulty index.

The second aspect is that there are three features of the item wording and figure that may mislead students. The first feature is the use of the term "uncharged" (first step of reasoning) that may mislead students to think that there is absolutely no charge in the conductor, and/or to select the most frequent error. Note that in the Spanish version we use the same term. The second is regarding all choices of the item. We understand that the item asks for the net forces on the point charges q and Q, and thus the choices have to talk about these forces. However, mentioning exclusively the forces on these two objects does not help students to think in terms of the entire system, which is related to the second step of reasoning. This may also mislead students to select the most frequent error. The third feature is that the figure does not show the inner and outer surfaces of the sphere, related to the third step of reasoning.

C. Analysis of item 20

Figure 2 shows the figure and options of item 20. The question is as follows:

20. A positively-charged proton is first placed at rest at position I and then later at position II in a region whose electric potential (voltage) is described by the equipotential lines. Which set of arrows on the left below best describes the relative magnitudes and directions of the electric force exerted on the proton when at position I and II?
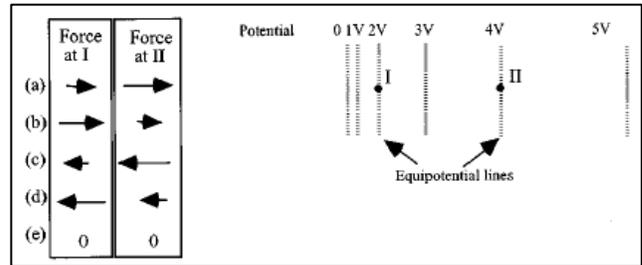
FIG 2. Figure and options of item 20 of the CSEM.

Item 20 is another of the most complicated items. As shown in Table II, the correct answer percentages for the AMU and MXU students are very low (option D, 25% and 19% respectively). Two incorrect options in both populations have percentages slightly higher than those of the correct answer (options B and C). B is an option with forces opposite to the correct directions in both locations but with correct relative magnitudes. This option indicates that students incorrectly think that an increase in potential determines direction [4]. Moreover, C is an option with correct force directions in both locations but with incorrect relative magnitudes. This option seems to indicate that students are associating large distances between equipotential lines with stronger fields or forces [4]. Also, this answer could indicate that students are thinking that field or force is larger where potential is larger [9]. To analyze the problematic aspects of this item, we do the same analyses as for item 14.

The main problematic aspect that we noticed in this item is that its reasoning process (shown in Table IV) has three characteristics. The first is that, as in item 14, the reasoning process is very elaborate and has five steps of reasoning. The second characteristic is that the process involves many variables: potential, potential difference, electric field, electric force, direction and relative magnitudes of electric fields and forces. The third is that the process involves many relationships, using three different equations.

These three characteristics seem to explain why in both populations the correct answer proportions are very low and that there are two incorrect options (B and C) that have percentages slightly higher than that of the correct answer. In our previously mentioned study [3], we noted that an item of a mechanical waves test with exactly these three characteristics also showed a very low difficulty index.

TABLE IV. Required reasoning process to answer item 20.

Step	Description of the reasoning
1	Realize that, to answer the force question, they have to think about the electric field since $\vec{F} = q\vec{E}$ and, since the charge is positive, then the directions and relative magnitudes of the forces exerted on the proton at positions I & II are the same as those of the field at those positions.
2	Think in terms of electric field and determine qualitatively the directions and relative magnitudes of the electric fields at the two positions.
3	Use the equation $\Delta V_{1-2} = -\int_1^2 \vec{E} \cdot \overrightarrow{ds}$ to determine that a change of electric potential determines direction, e.g., going to the right increases the potential, i.e., ΔV_{1-2} is positive and to make the integral $-\int_1^2 \vec{E} \cdot \overrightarrow{ds}$ positive, the dot product should be negative, thus realize the electric field is to the left at the two positions.
4	Use the same equation or the inverse equation $E = -\Delta V / \Delta x$ to determine that, for the same potential difference, a shorter distance between the equipotential lines means a stronger field, thus that the greater field is at position I.
5	Conclude that, because the electric field is greater at position I, then the greater force is exerted on the proton at that position.

In this item, there is one feature of the item wording that could be problematic. The item says: "…*region whose electric potential (voltage) is described*". We argue that electric potential is not voltage; whereas electric potential difference is voltage. However, we have no evidence that this may cause problems to students.

V. DISCUSSION AND CONCLUSION

In this article we analyzed in detail two items of the CSEM (items 14 & 20) with the most critical index values in two populations of different institutions. In both items, we found that the main problematic aspect is that the required reasoning process to answer them is quite elaborate, and in item 14 we also found some misleading features of the wording and figure. These problems with the items may have an effect on the test as a whole. The CSEM was designed to be used in a 50-minute period; since the test has 32 items, students have approximately one and half minute to answer each item. The elaborate procedures of these two items seem to suggest that they would be better placed in an assessment instrument with open-ended problems, with each of the problems having several questions guiding the students.

The previously analyses also suggests that the CSEM should be refined as other assessments have been [3]. Each researcher should consider what to do with these two items; however, we recommend making changes concerning them in a next version of the test. In the case of item 14, the change could contemplate the following possible modifications: an item that evaluates the subject of induced charge with a less elaborate reasoning process, the use of "neutral" instead of "uncharged", a more accurate figure, and to describe all the forces in the system in the choices. This new version of the item would have to be validated. Planinic [5] found through Rasch model that item 14 was problematic and in her conclusions she recommends to delete it from the test or to modify it, as we also do. In the case of item 20, it is important to note that items 18 and 19 separately assess the two concepts evaluated simultaneously in item 20 (electric field and force in equipotential regions). Since the reasoning process for item 20 is very elaborate, it might be a good idea not to include this item in the next version of the test. It is very important to mention that with these changes researchers will have to make adjustments to compare the data of the new version with the data of the original version.

There are three important differences between Planinic's [5] and our study. The first difference is that she uses data from Croatian students. In contrast, in our study we use data from American and Mexican students to reduce the effect of the specific nature of the population. The second is that in our study we analyze in detail the possible sources of problems of item 14. Finally, the third difference is that we present specific recommendations for possible changes on item 14. Note also that Planinic does not identify item 20 as a problematic item. It is interesting that, through different analyses such as the Rasch model and the statistical tests focused on each item that we performed, we arrive to a conclusion similar to Planinic's about item 14.

The analyses and recommendations of this article are important because the CSEM is widely used to assess the effectiveness of new research-based curricula. These analyses and recommendations may be useful for researchers using the test as an assessment tool and for a possible future version of the test. In a future article we will continue analyzing other items of the test with critical index values and other problematics aspects of the test.

[1] E. Redish, Am. J. Phys. **67**, 562 (1999).

[2] L. Ding, R. Chabay, B. Sherwood & R. Beichner, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).

[3] P. Barniol & G. Zavala, Phys. Rev. Phys. Educ. Res. **12***, 010107 (2016).

[4] D. Maloney, T. L. O'Kuma, C. J. Hieggelke & A. Van Heuvelen, Am. J. Phys. **69**, S12 (2001).

[5] M. Planinic, GIREP conference 2006, **20** (2006).

[6] R. A. Serway & J.W. Jewett, *Physics for Scientists and Engineers (8th edition)* (Cengage, USA, 2010).

[7] H. Young & R. Freedman, *University physics with modern physics (13th edition)* (Pearson, Mexico, 2013).

[8] C. Henderson, Phys. Teach. **40**, 542 (2002).

[9] D. E. Meltzer, AIP Conf. Proc. **883**, 177 (2007).