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).