Comparing exploratory factor models of the Brief Electricity and Magnetism Assessment and the Conceptual Survey of Electricity and Magnetism

Philip Eaton,[*] Barrett Frank, Keith Johnson, and Shannon Willoughby

Department of Physics, Montana State University, 1325-1399 S 6th Avenue,
Bozeman, Montana 59715, USA

While numerous studies have analyzed the conceptions probed by the Force Concept Inventory (FCI), assessments dedicated to electricity and magnetism lack similar analyses. This paper investigated the conceptions explored by the Brief Electricity and Magnetism Assessment (BEMA) and the Conceptual Survey of Electricity and Magnetism (CSEM) using exploratory factor analysis techniques. Exploratory factor analysis (EFA) was performed on both assessments using 5368 and 4941 postinstruction student responses for the BEMA and CSEM, respectively. A 6-factor EFA generated model was found for the CSEM, and was fit against another sample of 4964 student responses using confirmatory factor analysis to supply evidence for the possible generalizability of the model. The 5-factor EFA generated model for the BEMA could not be fit against another sample when trying to check for generalizability. The EFA generated factor models for the BEMA and CSEM were then compared and found to be similar in conceptual content, with the exception of one or two factors. Thus, from a factor analysis perspective, the BEMA and CSEM were found to be similar in conceptual content as revealed by student responses. With a better understanding of these electricity and magnetism assessments, future research into this domain of physics will then be able to make stronger conclusions based on students' results within these assessments.

I. INTRODUCTION

The Brief Electricity and Magnetism Assessment (BEMA) and the Conceptual Survey of Electricity and Magnetism (CSEM) were presented to the physics education research community in 2006 and 2001, respectively [1,2]. Since their validation papers, these assessments have become common tools for evaluating students' conceptual knowledge of electricity and magnetism. However, these assessments have not undergone significant psychometric evaluation. Both assessments have been found to be valid and reliable, see Refs. [1,2], but little has been done since then to further understand the nature of these assessments. A deeper understanding of the statistical properties for these instruments will enable future studies that use these assessments to come to more informed conclusions.

This study seeks to uncover and explore prominent student response patterns (factor structures) which will aid in investigating what conceptions students may be using when taking each of these assessments. Specifically, we will investigate the factor structure of both the BEMA and CSEM. These factor structures will be tested in a future study, which will focus on identifying statistically valid and rigorous factors for each of these assessments. Previous statistical studies of note for each assessment are their respective validation papers (see Refs. [1,2]), a gender grouped differential item functioning analysis of the CSEM [3], a classical test theory comparison of both assessments [4], and a classical test theory combined with item response theory comparison of both assessments [5].

The comparisons made in Refs. [4,5] demonstrate that these assessments are similar in difficulty from a classical test theory and an item response theory point of view. However, both of these techniques make the assumption of unidimensionality (i.e., a single conceptual factor) when examining assessments at an item level, and thus do not give information on the multiple conceptions probed. The most significant difference between these assessments is the presence of circuits specific questions on the BEMA, while the CSEM contains none.

The intention of this study is to *explore* and compare the factor structure of these assessments. This study does not intend to verify that students are answering the questions in the manner intended by the creators of the BEMA and CSEM, nor are we confirming the factors found as a result of our exploration. We only seek to compare how students are responding to each of these assessments. Understanding how the factor structures of each assessment compare to one another will help instructors and researchers in

[*]philip.eaton@montana.edu

identifying the most appropriate assessment to use in their classes or studies. Further, the proposed models found here can be used in future studies to help develop expert-proposed models.

To investigate if postinstruction student-driven models generate factors which probe concepts of electricity and magnetism, exploratory factor analysis will be applied to each assessment. Exploratory factor analysis (EFA) investigates latent traits (i.e., factors) within the assessments that help explain the variance in the student response data. However, since EFA is not suited to validate factor models, the exploratory-driven factor structures will need to be validated in a future study using confirmatory factor analysis (CFA) and an alternative set of data. A future study will consist of using the models found within to help inform the construction or modification of expert-proposed models for the BEMA and CSEM.

Since the BEMA and CSEM were developed for first-time introductory electricity and magnetism classes, it is expected that the conceptions probed by the factors will be similar, with the exception of circuits. So, this study sought to address the following two research questions:

RQ 1: What are the student-driven factor structures generated through exploratory factor analysis for the CSEM and BEMA?

RQ 2: How do the two resulting models for the CSEM and BEMA compare, and do they assess the same or similar conceptions?

Previously, it has been suggested by Maloney *et al.* that the CSEM is a unidimensional assessment [2]. However, without knowledge of the exact methodology used we cannot comment on how the researchers arrived at this conclusion. It should be noted that Maloney *et al.* applied a principal component analysis to the CSEM [2], whereas the analysis performed in this study uses principle axis factoring. These two factor extraction methods make different statistical assumptions, thus it is not unreasonable to assume that they will generate slightly different results. The exact differences between these methods will not be discussed in this paper, but are addressed in multiple references, such as Ref. [6].

The article is structured into seven distinct sections, including the introduction, Sec. I. Section II discusses aspects of the data used in this study, and Sec. III presents the statistical methodology used to answer the research questions. The results of the analysis, as well as a comparison of these results for both assessments, are presented in Sec. IV. Lastly, the limitations of this study, the study's conclusions, and intended future works are detailed in Secs. V–VII, respectively.

II. DATA

The data used in this study were supplied by Physport, an online support website for physics instructors at all levels of education. Physport contains a collection of conceptual inventories with varying amounts of student response data

TABLE I. Test statistics for the EFA and CFA CSEM datasets and for the full BEMA dataset.

	EFA CSEM	CFA CSEM	BEMA
N	4941	4964	5368
Average	0.441	0.442	0.456
St. Dev.	0.187	0.185	0.173
Skew	0.496	0.474	0.071
Kurtosis	−0.380	−0.424	−0.686
Cronbach's α (KR-20)	0.83	0.82	0.82
KR-21	0.83	0.82	0.82

for each [7]. The data for the BEMA and CSEM consisted of post-test student responses from a mixture of algebra- and calculus-based introductory physics classes, all using potentially different teaching strategies (i.e., traditional lecturing, inclusion of active learning, etc.). Upon receiving the data from Physport, any student responses with even a single unanswered question were removed from the dataset. Each instrument was scored using the specified grading criteria offered for each assessment. The CSEM's data consisted of 9905 student responses, of which 4941 were used for EFA and 4964 were used for confirmatory factor analysis. To generate two samples for the CSEM the students were given random identification numbers, and the students with even identification numbers were used for the EFA and the odd students for confirmatory factor analysis (CFA). In total, the BEMA had 5368 student responses, which could not be split into two groups due to fit issues explained in Sec. IV.

The test statistics for each of these groups can be found in Table I. As can be seen, the two groups of data for the CSEM behaved similarly after separating them from one another, with mean total score values found to not be significantly different (two-tailed *t* test: $p = 0.9245$). Both assessments have relatively low averages ($<50\%$), indicating that these assessments tend to be rather difficult for the students in an introductory physics class, even after instruction. Results like this are well understood and have been discussed in other studies [1,2,5,8]. Further, the Cronbach's alpha and the KR-21 have been calculated for each of the samples, see Table I [9]. It should be noted that the Cronbach's alpha reduces to the KR-20 in the limit that the data is dichotomous. Comparing these results to the validation papers for each assessment gives good agreement with an alpha greater than 0.8 for the CSEM and BEMA each. Thus, it can be concluded that when the students answered items on these assessments they did so in a consistent and reliable manner. It should be noted here that Cronbach's alpha is not a test statistic for unidimensionality, it is simply a measure of the internal consistency of an assessment [9].

III. METHODOLOGY

Exploratory factor analysis, explained in detail in the following section, was employed to investigate the

groupings of questions which form from correlations between questions within the student responses. Using the factor extraction methods discussed below, factor models were constructed to suggest how items (questions) on the assessments group together using item correlations. Each factor within these models is ideally probing one specific conception of the students. These models were then fit against the set of data which produced the EFA generated model to check for goodness of fit; this was done using CFA. To support the claim that these models are not unique to the sample which produced the models, CFA on independent samples was performed when possible. For the CSEM, the EFA generated model could be fit to the other, independent, half of the data. However, this could not be done for the BEMA, which is discussed in detail within the results section below. If the models were found to have acceptable goodness of fits with the alternate sets of data, then the models were said to be adequate representations of the correlation groupings of the items on the assessment.

A factor is taken to represent some underlying concept a student employs when responding to the set of questions which make up the factor. The qualitative interpretations of the factors presented in this study are one of many possible logically consistent interpretations of the factors. The ambiguity in these interpretations is an aspect of offering an expertlike explanation for possibly novicelike thought processes. This inevitably leads to many possible explanations for groupings sourced from nonexpert thinkers. Student interviews should be used to lend support for the interpretations of each factor, and is suggested as a future study.

After identifying what physical idea the factors are probing, comparisons of the concepts investigated by each of the assessments can be made. This will lend insight into the similarities and differences of the content explored by each instrument according to the student samples used in the analysis.

The following two subsections are brief explanations for how EFA and CFA were performed, and the tools that were used in this analysis.

A. Exploratory factor analysis

Exploratory factor analysis is a statistical tool that groups items on an assessment together based on item correlations generated from student responses. Since the data in this study was dichotomously graded (0–incorrect, 1–correct) a tetrachoric correlation was used to generate the correlation matrix [10]. This kind of analysis allows researchers to investigate two intermixed ideas: (i) how student responses form factors based on the correlations between the student responses to questions, and (ii) what these factors might represent according to an expert interpretation (i.e., specific physical concepts). One of the key aspects of EFA is that the number of factors used in the generated models is

selected by the researchers using one of many proposed methods (see Refs. [11–13] for a few examples).

Mathematically, a factor is a numerical vector of item loading values, where item loadings (often called "loadings") are related to the amount of variance for an item that is explained by the factor being considered [14]. Loadings can take on values between -1 and 1, with numbers closer to ± 1 meaning the items are better explained by the factor when compared to those closer to 0.

Conceptually, factors are assumed to represent the conceptions or latent traits probed by the assessment being analyzed. Often, identifying the conception that a factor is representing is based on the items which load best onto a factor. Thus, a factor must contain multiple items to best assess the concept a factor may be representing. A common criteria within factor analysis is that each factor should contain at minimum 3 items [14].

This study is not concerned with how adequately a certain conception is probed by the assessments. Instead, the interest is specifically on which conceptions are explored by the assessments under consideration. As a result, a smaller than normal item loading cutoff value of 0.2 was used, with minimal exceptions, and factors were allowed to have less than 3 items. This means factors found within this analysis will not be able to assess how well a student understands the conception characterizing the factor. For quantifiable scales to be developed, more stringent cutoff values will need to be used. There are many suggestions for what a minimum loading should be to place an item onto a specific factor [6].

To generate the factor models investigated in this study, the principle axis methodology, see Ref. [15], was implemented using the R function "fa" [16] with an oblimin rotation [15,17]. Rotations to the extracted factors, like the oblimin rotation, are performed to improve the fit of the factors to the data [15,17]. An oblimin rotation is nonorthogonal, meaning the rotation allows the resulting factors to have angles between them that are less than 90° in parameter space. The cosine of these angles can be interpreted as the correlation between factors. A nonorthogonal rotation was used versus an orthogonal rotation due to two reasons. First, it is likely that the conceptions which caused the factor to form will be related to one another. Second, if the resulting factor structure is naturally orthogonal then a nonorthogonal rotation will be able to capture this behavior (i.e., nonorthogonal rotations are more general) [15,17].

Prior to performing EFA, the number of factors to be used must be identified. There are many methods suggested for identifying the number of factors that could be extracted from the data [11–13]. As it stands, the community as a whole does not have an agreed upon method to identify the proper number of factors to use when generating an EFA model. This study opted for using an approach that combines many of the proposed processes currently used.

Utilizing an algorithmic model building method (discussed in the remainder of the section) and fitting the created models to the data that generated them using CFA, the model that best describes the data was found.

This algorithm begins by selecting cutoff values for (i) the item loadings, (ii) the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy values, and (iii) the item communalities. KMO values can be calculated for each of the items on an assessment, and are a measure of the sample's ability to place the items onto factors [18]. These values will be between 0 and 1, with numbers closer to 1 meaning the item can more confidently placed onto the factors. It is suggested that items with KMO values above 0.8 be kept, with 0.6 being the lower limit of the possible cutoff values [6]. Item communality is a numerical measure of how much an item's variance is being captured by the factor model [14]. Communalities between 0.25 and 0.4 have been suggested as acceptable cutoff values, with ideal communalities being 0.7 or above [6].

Generally, the stricter these cutoff values the better fit the model has with the items that remained. So, choosing smaller cutoff values for item communality results in models with potentially poorer fits compared to ones generated using stricter cutoff values. As a result, the models generated using this algorithm will have varying degrees of goodness-of-fits with the data depending on the strictness of the cutoff values initially selected. This must be balanced with the fact that stricter cutoff values can result in fewer items being retained within the model. So, cutoff values should be selected such that the generated models fit the data well and retain an acceptable number of items from the assessment.

Once these cutoff values have been selected, the number of factors to be extracted is chosen. This number should not exceed the maximum number of factors dictated by parallel analysis, which models the random guessing of students [12]. This model, called the baseline model, uses a correlation matrix generated by selecting uniformly, randomly sampled values between -1 and 1 for the entries of the matrix. Eigenvalues are extracted from this random guessing correlation matrix, and are interpreted as the amount of variance that belongs to noise (i.e., random guessing). These eigenvalues are placed in descending order and compared to the eigenvalues of the data's correlation matrix; also placed into descending order. The data's eigenvalues that are greater than the random guessing eigenvalues are said to be above the noise and are capable of being extracted. Thus, parallel analysis places an upper limit on the number of factors an EFA model can possess.

Once the cutoff values and the number of factors to be extracted have been selected the algorithm is performed using the following steps. If an item is removed from the analysis at any point in these steps, then the algorithm restarts at step 1, unless otherwise indicated. The steps are as follows:

(1) Calculate the KMO values. If any items have a KMO value below the cutoff value, then remove the item with the smallest KMO value and repeat this step.

(2) Use the Bartlett test of sphericity to check that the correlation matrix can be factored. If it cannot, then factor analysis cannot be performed.

(3) Calculate the EFA model using principle axis factoring with the specified number of factors.

(4) Calculate the communalities. If any items have a communality below the cutoff value, then remove the item with the minimum communality and restart at step 1.

(5) Calculate the item loadings. If any items have maximum loadings across all of the factors that fall below the cutoff value, then remove the item with the smallest maximum loading and go back to step 1.

(6) Create a factor model by placing each item onto the factor that contains the item's largest loading value. If any items load equally onto more than one factor, then allow them to cross-load onto multiple factors. Fit this model to the original data using CFA, making modifications as is suggested by CFA criteria (see Ref. [14]), and finally extract a parsimonious fit statistic [in this study the Akaike information criterion (AIC) was used, but other parsimonious fit statistics would be acceptable]. How CFA is performed and the meaning of a parsimonious index is discussed in the following subsection.

(7) Change the number of factors and repeat steps 1–6. (Models with various numbers of factors should be generated when the process is completed.)

(8) Plot the parsimonious fit index (i.e., the AIC) against number of factors. The model with the (local) minimum index is the preferred model, and is the one extracted by this algorithm.

It should be noted that due to the nature of EFA, items will occasionally load onto a factor with a small loading value relative to the other items on the factor. This is due to the mandate in EFA estimation methods to use all of the items given in the data. These items are generally ignored when discussing the qualitative meaning of the factors.

B. Confirmatory factor analysis

Confirmatory factor analysis is a model driven analysis that tests a given factor model's ability to recreate the correlation matrix of the data which the model is fit against. The model, which can be found through EFA or can be built from an expert's consideration of the items, is a prescription of which items load onto each factor. This is different from EFA since there are no cross loadings of items, unless the model is created with them initially. A model found in EFA will inherently have loading values on all of the factors, and are ideally very small compared to the dominant loading value. A general rule of thumb is if a loading value is smaller than the dominant one by 0.3 it can be ignored, otherwise the

item's cross loading should be kept in the model [19]. These cross loadings in EFA may artificially improve the fit of the model to the data. Thus, when the EFA model is moved to a CFA model, and the small cross loadings are removed, the goodness-of-fit will generally change. It is this reason that step 6 was incorporated into the algorithm used in this study's EFA model selection method.

The model used in CFA is fit to the data by estimating the loadings of the items to their factors given the correlation matrix of the data. This can be done in a number of different ways (e.g., maximum likelihood methods) that will not be detailed in this study; explanations of these methods can be found in Ref. [14] and the references cited within. To perform these estimations the R software "lavaan" was used [20].

After the model has been fit to the data, goodness-of-fit statistics can be calculated. For example, there are the confirmatory fit index (CFI) and the Tucker-Lewis index (TLI) that compare the fit of the proposed model to a baseline model (a model that makes no assumptions about the fits of the items). These fit statistics are said to be acceptable if they are above 0.9 [14]. There is also the standardized root mean square residual (SRMSR) and the root mean square of the error of approximation (RMSEA), both of which should be below 0.08 [14]. Lastly, there are parsimonious fit indexes [like the Akaike information criterion or the Bayesian information criterion (BIC)] which give a goodness-of-fit measure that is related to the likelihood of the model and an added penalty based on the number of free parameters in a model. These statistics do not have a range of acceptable values, but instead are used to compare models of similar goodness-of-fit to identify the preferred model. The preferred model is generally the one with the smallest value for its parsimonious fit indexes. More information on these fit indexes, as well as on how CFA is performed, can be found in Refs. [14,21].

IV. RESULTS AND DISCUSSION

Before investigating the multidimensionality of the assessment, the potential unidimensionality was checked. When performing a 1-factor EFA for the CSEM and BEMA the fit statistics were found to be unsatisfactory (CSEM TLI $= 0.786$; BEMA TLI $= 0.841$). This indicates that a unidimensional model is inadequate for representing these assessments within a factor analysis framework.

The following is a presentation and discussion of multidimensional results for the CSEM and BEMA using EFA. Readers are encouraged to have a copy of the CSEM and BEMA on hand to reference questions being discussed when the qualitative meaning of each factor is presented below.

A. EFA of the CSEM

1. CSEM 6-factor EFA model

For the EFA of the CSEM the following cutoff values were used: item loadings ≥ 0.20, KMO values ≥ 0.6, and

item communalities ≥ 0.15. As a result, the following items had to be removed from the analysis: 1, 2, 14, 22, 28, 30, and 32. These items were dropped due to poor item communalities. However, it is worth noting that the KMO values were all above 0.80, with the exception of 0.73 for item 14. In an attempt to understand why these items did not load onto the EFA model for the CSEM, the classical item difficulties were calculated for all of the items. However, no consistent explanations for why an item did or did not load could be determined after investigating these indexes. A similar analysis was performed for the BEMA, and again no consistent explanation for loading and nonloading items could be determined.

A parallel analysis of half of the CSEM's data revealed that at most 12 factors could be extracted. However, once items were dropped due to communality issues, this value dropped to 10. The AIC values for the EFA models ranging from 2 to 8 factors can be found in Fig. 1. The 9- and 10-factor models are not presented since they contained multiple factors with only 2 items or fewer loading on to them, and thus were removed from consideration for model selection.

It can be seen that the 2-, 4-, and 6-factor models are the local minima of the AIC versus number of factors plotted for the CSEM; see Fig. 1. Thus, they are the preferred EFA models that should be investigated further. The 6-factor model was selected since it contains more information (i.e., questions and factors or concepts) about the assessment compared to the other models. The item loadings for the 6-factor model can be found in Table II. In the validation paper for the CSEM, see Ref. [2], the creators offer a suggestion on the concepts each item probe. Many of the
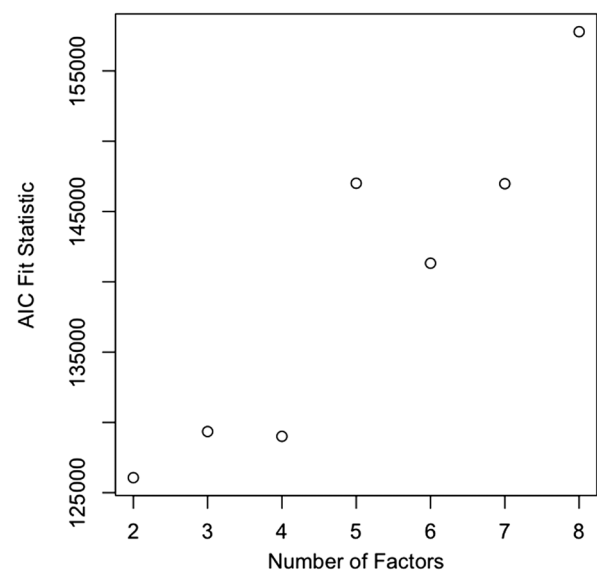

FIG. 1. The AIC values for the factor models that resulted from the criteria laid out in Sec. III A for the CSEM used in CFA with the EFA half of the CSEM dataset. As can be seen, the 6 factor model is the locally preferred factor model.

TABLE II.   CSEM 6 factor EFA model. The following items had to be dropped due to their communality values falling below the 0.1 cutoff value: 1, 2, 14, 22, 28, 30, and 32. All loading values above 0.2 were retained in the presented model. Within this table "Q" indicated the questions number and "L" the associated loading value.

| Factor 1 | | Factor 2 | | Factor 3 | | Factor 4 | | Factor 5 | | Factor 6 | |
| Newton's 3rd law | | $\vec{F}_E = q\vec{E}$ + Superpos. | | $\vec{B}$ by I | | $\vec{F}_B = q\vec{v} \times \vec{B}$ | | $\vec{E} = -\nabla V$ | | q in fields | |
Q	L	Q	L	Q	L	Q	L	Q	L	Q	L
4	0.826	8	0.521	23	0.669	21	0.495	20	0.401	11	0.343
5	0.628	6	0.473	26	0.624	27	0.363	18	0.330	15	0.312
7	0.547	9	0.411			25	0.350	19	0.218	10	0.247
24	0.310	12	0.409			29	0.315			19	0.213
		3	0.405			31	0.289				
		17	0.256			16	0.205				
		16	0.207			13	0.202				

factors found for this 6-factor model align well with the creators' suggestions.

The first of the factors for the 6-factor model of the CSEM contained items 4, 5, 7, and 24 and pertains to Newton's 3rd law. As suggested by the creators of the CSEM, all of these items require an understanding of action or reaction pairs in the context of electricity and magnetism. Items 4, 5, and 7 all ask about forces acting on positive and negative point charges. These questions could be answered using Coulomb's law, however item 24 could not. Item 24 asks about the force between two current carrying wires. The inclusion of item 24 on this factor implies that it is probing Newton's 3rd law over Coulomb's law.

Items 3, 6, 8, 9, 12, 16, and 17 all appear on the second factor and probe the idea of determining the electric force on a charge given a charge distribution or an external electric field. Items 3, 6, 8, and 9 refer to point charge configurations and their associated electric field or force at a specified location and/or charge, with items 8 and 9 requiring knowledge about superposition of vector fields. Item 12 asks about the force on a charge placed into an electric field, and items 16 and 17 ask about the force on a charge placed into an electric potential. However, since items 16 and 17 loaded quite poorly onto this factor, they will not be used to offer a conceptual interpretation for this factor. Since the largest loading items on this factor all require knowledge that the electric force is equal to the charge feeling the electric force times the electric field at the point of the charge, with some questions needing a few more conceptual steps than the others. Ultimately, the common idea probed by all of these items is $\vec{F}_E = q\vec{E}$.

The third factor of the CSEM's 6-factor model contains only two items, which investigate student understanding of magnetic field generation from current carrying wires. Items 23 and 26 both deal with the magnetic field produced by current carrying wires. Item 23 contains two current carrying wires, and will thus need superposition to answer, whereas item 26 only has one wire.

Related to the third factor, the fourth factor is believed to probes students' understanding of the magnetic force on charged particles. Items 21, 25, 27, and 31 all explicitly require the use on the right-hand rule for the magnetic force on a charged particle. Item 29 inquires about Faraday's law of induction, which can be related to the magnetic force on charged particles. Items 13 and 16 appear to be out of place on this factor since they probe electrostatic conceptions. These two items happen to have poor loading values compared to the other items, see Table II, which may result from reducing the number of factors from 8 to 6. It could be that this factor is the results of students' understanding of the Lorentz force law, which combines the electric and magnetic forces. Though, with a factor in this model dedicated to examining electric forces on a charged particles due to electric fields, this interpretation does not appear to be consistent with the factor model. Thus, using the better loading items (21, 25, 27, 29, and 31) it can be concluded that this factor, for the most part, inquires about magnetic force acting on a charged particle, which requires the use of the right-hand rule to one degree or another.

Factor five contains 3 items (18, 19, and 20) which ask about the electric force or field acting on a charged particle placed into an electric potential difference. Item 18 asks for a ranking of the electric field of three plots of electric potentials, and item 19 asks about the direction of the electric force exerted on a positively charged particle for one of the cases discussed in item 18. Lastly, item 20 inquires about the direction and magnitude of the electric force acting on a proton that is in an electric potential difference whose plot is similar to the second one discussed in item 18. This factor probes students' ability to infer vector fields from the scalar potential.

Finally, factor six contains items 10, 11, 15, and 19, which are concerned with the relations between electric fields, electric forces, and electric potentials. For example, item 10 asks about the resulting motion of a charged particle placed into an electric field, and item 11 asks about how the electric potential energy changes for the charge in item 10 as it moves. Similarly, item 15 asks for the direction of the electric force for a negative point charge placed into an electric field, and item 19 asks for the direction of the

electric force exerted on a positive charge when placed into a potential gradient. This factor appears to cover some of the same ideas as factors 2 and 5. However, Table V reveals that factors 2 and 6 as well as 5 and 6 are not well correlated. Thus, this factor is measuring something the other two factors did not cover. So, this factor seems to probe the interactions between charged particles and the electrostatic fields around them.

Exploratory factor analysis does not have a medium that allows for a discussion for the conceptual reason governing why questions were removed from the analysis. Thus, the rationale offered for why any single item was removed is purely speculation on the part of the researcher. For example, when considering the removed CSEM items no overarching conceptual theme could be seen amongst more than two or three of the items. This is contrasted by the interpretation of factors, where a more consistent rationale is the derivative of multiple, well-loaded items.

B. EFA of the BEMA

In an attempt to repeat the methodology applied to the CSEM (i.e., splitting the data into two separate samples for EFA and CFA to be applied separately) the BEMA data was initially split into two groups. However, fitting the EFA generated model against the other half of the data, using the R software lavaan, caused a divergence in the fit. After the groups were merged and the full sample was used, the EFA algorithm generated acceptable models with good fit statistics for CFA. It is not clear why splitting the data in half caused issues, and indicates that further study of the BEMA should be done using a different set of data. As a result, since a CFA sample group was not able to be formed, the EFA model generated could not be tested against an alternate, independent sample. This should be done in a future study to check if the generated model is unique to the sample or if it is potentially generalizable.

Using the following cutoff values: item loadings ≥ 0.175, KMO values ≥ 0.6, and item communalities ≥ 0.10 with the full BEMA sample, it was found that a 5-factor model was the preferred model for the data, shown by the AIC fit statistic in Fig. 2. As a result of the cutoff values selected the following items were removed from the model: 7, 8, 9, 10, 11, 14, 17, 18, 19, 27, and 28/29. Items 9 and 18 were removed due to poor KMO values, and the others were removed due to poor item communality values. The factors can be found in Table III, and the meaning of the factors will be discussed for the remainder of this section.

The first factor for the 5-factor BEMA model contains items 6, 12, 20, 23, 25, 26, and 30 and is interpreted as inquiring about the Lorentz force. Items 20, 23, 26, and 30 ask about situations where charged particles move through a magnetic field and feel a resulting magnetic force. Item 25 contains a figure with two current carrying wires and asks about the force on one wire due to the other. Items 6 and 12
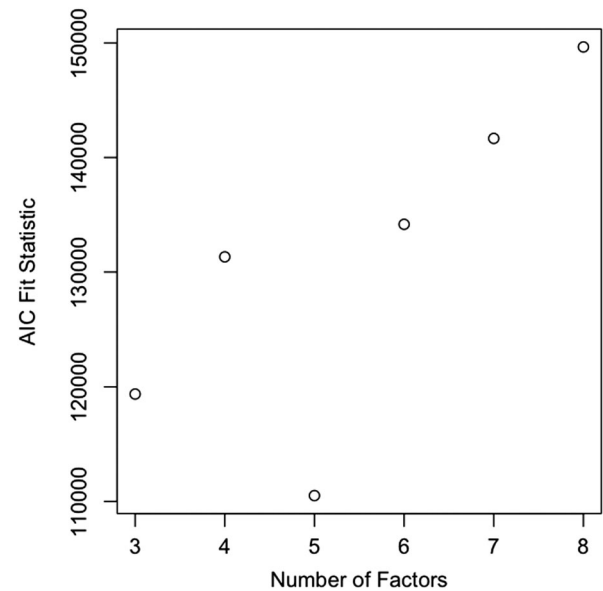

FIG. 2. The AIC values for the factor models that resulted from the criteria laid out in Sec. III A for the BEMA used in CFA with the full BEMA dataset. As can be seen, the 5 factor model is the preferred factor model.

are not as well loaded onto this factor compared to the other items. Item 6 asks about an electric field that is acting on an electron and item 12 inquires about the direction of an electric field acting on electrons while they are traveling through a light bulb that is hooked up to a battery. Since the items contained by this factor appear to be about the electric and magnetic forces acting on charged particles, this factor can be inferred to probe the Lorentz force, see Table III.

Items 21, 22, and 24, which make up the second factor for this model, pertain to the magnetic field generated by a bar magnet at two locations (items 21 and 22) and about the magnetic field generated between two loops of current carrying wire (item 24). This factor has been labeled magnetic fields since the most prominently loaded items for this factor are concerned with magnetic field generation. Item 13, the poorest loading item on this factor, asks for the current which results from a charged capacitor decaying after it is connected to an ammeter. This item is thought to exist on this factor due to the nature of EFA utilizing all items it is given within the generated model, and not due to conceptual content.

Factor three contains items 15 and 16, which touch on the electric potential difference between two points in a uniform electric field. It is interesting that item 14 did not load onto this factor considering it also asks about the electric potential difference between two points in a uniform electric field using the same figure as items 15 and 16. Item 14's removal as a result of small item communality suggests that this question had little in common with the other items on this assessment. This lack of commonality is not obvious to the authors, and student interviews for question 14 may help elucidate its removal from this factor.

TABLE III. BEMA 5 factor EFA model. Items 9 and 18 were dropped due to poor KMO values, and the following items were dropped due to their communality values falling below the 0.1 cutoff value: 7, 8, 10, 11, 14, 17, 19, 27, 28/29, and 30. All loadings above 0.15 were used for this model. Here "Q" indicates the questions number and "L" the associated loading value.

Factor 1		Factor 2		Factor 3		Factor 4		Factor 5	
Lorentz force		Mag. field		ΔV given $\vec{E}$		Coulomb's law		Elec. field	
Q	L	Q	L	Q	L	Q	L	Q	L
30	0.486	21	0.703	15	0.907	2	0.794	4	0.690
26	0.477	22	0.699	16	0.443	1	0.543	5	0.437
20	0.412	24	0.234			3	0.178	6	0.223
23	0.378	13	0.220						
25	0.364								
12	0.318								
6	0.267								

The ideas concerning Coulomb's law explain the grouping of items 1, 2, and 3 for the fourth factor of the BEMA. These items examine the electric force between two particles and how it changes relative to the original configuration, when aspects of the particles are changed. Item 1 increases the charge of a particle and asks about the resulting force on the unchanged particle, item 2 asks about the force on the changed particle, and item 3 moves the particles farther apart and asks about the resulting force acting on the changed particle. As a result, this factor has been labeled Coulomb's law.

Lastly, factor five investigates electric field or force generation. Items 4 and 5 provide a charge distribution and ask about the direction of the resulting electric field generated at two points in space. Item 6 has an electron move through a region that contains an electric field and shows how the path of the electron changes in that region. Based on this information, the direction of the electric field in the unknown region of space is inquired about. It should be noted that the Coulomb's law factor asks explicitly about the force on the charged particles, whereas this factor asks about the electric field. Knowledge of how the electric force and the electric field are related could be used by students to answer all of these questions, which is where this factor gets its qualitative interpretation.

C. Confirmatory factory analysis of the data-driven models

Performing CFA using the EFA generated model for the CSEM and the half of the data not used in the exploratory analysis resulted in acceptable fit statistics, see Table IV. This lends evidence that the EFA model is not sample specific and may be generalizable to the population the sample was obtained from. It is important to reiterate that the sample used in this study was taken from a broad spectrum of sources, and thus should be a good representation of all students taking an introductory electricity and magnetism course. For the BEMA, however, this analysis could not be performed. Since all of the data for the BEMA

was used in EFA, no independent sample was available at the time of this study. The CFA fit statistics presented for the BEMA in Table IV are using the data which generated the EFA model. The acceptable fit indexes suggest that the generated model adequately represents the samples but, without an independent sample, may not be generalizable. Examining the potential generalizability of this model is suggested as a future study.

D. CSEM and BEMA model comparisons

For each assessment, EFA was able to generate factor models which placed the items into groupings that generally probed a single expertlike conception as seen by the students. When interpreting these models, it is important to note that EFA is a data driven analysis that extracts the factor model that best fits the data used to generate it. From a correlation point of view this implies that the factors found represent how the sample as a whole grouped the questions together. These models will be compared to infer how two samples drawn from similar populations interact with the items on each assessment.

Since the BEMA and CSEM each probe similar conceptions with the exception of circuits, one would assume that any model for the BEMA would contain more factors than the one for the CSEM. However, since all of the circuit specific questions on the BEMA had to be removed when performing EFA, one would expect the models to have the

TABLE IV. CFA fit statistics for the two models found through EFA for the CSEM and BEMA. The CSEM was fit using the half of the data set aside for CFA, and the BEMA was fit back against the sample that originated the EFA model.

	CSEM 6 factors	BEMA 5 factors
CFI	0.939	0.959
TLI	0.928	0.949
SRMR	0.026	0.030
RMSEA	0.029	0.032
RMSEA 95% CI	0.031	0.034

same number of factors. Yet, the model for the CSEM contains one more factor than the BEMA, specifically comprehension of Newton's 3rd law. The remainder of this section compares the factors between these assessments.

The CSEM's third factor and the BEMA's second factor are very similar in that they both probe the generation of a magnetic field for different situations. The CSEM exclusively uses current carrying wires to generate the magnetic field, and the BEMA uses current carrying wires and a bar magnet. Though the generation of the magnetic field is considered for different situations within the two assessments, they each attempt to investigate student's understanding of how magnetic fields are generated in simple situations.

Similarly, the CSEM's fifth factor and the BEMA's third factor both consider the relationship between electric fields and electric potentials. The CSEM looks at how an electric field is generated from a potential difference, and the BEMA looks at the opposite situation, inferring a potential difference from a given electric field. Although the questions are superficially different, they are each probing the same content.

Both of the EFA generated models for the BEMA and the CSEM contain the concepts of electric and magnetic forces. The CSEM covers these conceptions using 3 factors (electric force, magnetic force, and Newton's 3rd law) and the BEMA does this with 2 factors (Lorentz force and Coulomb's law). Given the EFA models, the CSEM probes the electric and magnetic forces individually, whereas the BEMA probes them simultaneously through the Lorentz factor. Depending on how deeply one would like to probe electric and magnetic forces, the selection of assessment is important for these instruments. If the electric and magnetic forces independently are of interest to a researcher then the CSEM should be used.

The similarities between the EFA factors that were found for each assessment indicate the samples agree, with minor differences, that the BEMA and CSEM cover the same major concepts of electricity and magnetism in very similar manners, according to student responses. That a factor for the BEMA which probes circuits did not appear was unexpected. This could suggest that there may not be a coherent enough understanding of circuits within the student sample to be extracted through EFA.

TABLE V. The correlations between each of the factors for the CSEM after an oblimin rotation. Since the matrix is symmetric only the lower left corner is displayed.

	F1	F2	F3	F4	F5	F6
F1	1.00					
F2	0.40	1.00				
F3	0.38	0.65	1.00			
F4	0.39	0.35	0.39	1.00		
F5	0.20	0.34	0.29	0.29	1.00	
F6	0.35	0.44	0.44	0.36	0.27	1.00

TABLE VI. The correlations between each of the factors for the BEMA after an oblimin rotation. Since the matrix is symmetric only the lower left corner is displayed.

	F1	F2	F3	F4	F5
F1	1.00				
F2	0.58	1.00			
F3	0.46	0.49	1.00		
F4	0.42	0.35	0.33	1.00	
F5	0.55	0.63	0.49	0.41	1.00

Additionally, when considering the correlations between the factors for each model, all of the factors can be seen to be positively correlated with one another, as shown in Tables V and VI. This implies that the conceptions represented by each of these factors are distinct but are not independent. This is not surprising, since the ideas in physics generally play off of one another or share similar conceptual characteristics.

V. LIMITATIONS

The results of this study were generated using a data set that was a mixture of classes using different levels of mathematical sophistication and teaching techniques. Since EFA is data driven, and thus sample dependent, there could be differences between the factors for nonmixed data.

The qualitative interpretations supplied in this study are offered by the authors as possible meanings behind each of the factors found in the analysis. It is understood that there will be other interpretations that are equally logical and consistent with student thinking. This does not invalidate the factor structure found, which are backed up with statistical evidence, it is merely a reflection of the ambiguity that can enter when trying to supply a qualitative interpretation to the factors.

Lastly, since the BEMA's EFA generated model could not be fit to another independent student sample, the results cannot be assumed to be generalizable to the student population. Although the factors found made conceptual sense, it is still suggested that a future study be done to validate the model found in this study.

VI. CONCLUSIONS

As a result of this study potential factor structures for both the CSEM and BEMA have been generated. This gives instructors and researchers a clearer picture as to what concepts students are using when answering questions on these assessments. It was found that the CSEM is well modeled by a 6-factor model, which probes concepts like Newton's 3rd law, force due to an electric field, magnetic field generation by current carrying wires, magnetic forces on moving charges, and the relationship between an electric field and its associated electric potential, and the BEMA by a 5-factor model, which probes concepts like the Lorentz

force, magnetic field generation, the relation between an electric field and its associated electric potential, Coulomb's law, and electric fields due to point charges.

The factor structures for each assessment (see Tables II and III) examine similar domains of electricity and magnetism: electric forces and fields, magnetic forces and fields, and the relationship between voltage and the electric field. This is to say that students are using the domains of electricity and magnetism that an expert would expect when taking these assessments. However, the assessments may not approach certain concepts in identical manners. For example, the CSEM and BEMA investigate the relationship between voltage and electric field in an opposite manner (ΔV from a given $\vec{E}$ and vice versa). Ultimately, from a factor analysis perspective these assessments are effectively identical.

The correlations between the factors shown in Tables V and VI suggest that the conceptions measured by the factors are distinct but not independent. This lack of factor orthogonality is expected due to the presence of a single predominant eigenvalue within both assessments' correlations matrices.

VII. FUTURE WORK

Future work will move beyond exploring the manner in which students are responding to these instruments, and will focus on a formal validation of these assessments. This formal validation will utilize (i) confirmatory factor analysis of expert-proposed models, (ii) exploratory and confirmatory multitrait item response theory, and (iii) interview or free response data.

According to Ref. [2], when students take the CSEM as a pretest they do not appear to be randomly guessing. So, it would be informative to investigate the factor structure of pretest data for the BEMA and CSEM to see what questions group together. Also, further research into the effects that more homogeneous samples (i.e., algebra versus calculus) have on the resulting factor structures for each assessment is suggested.

ACKNOWLEDGMENTS

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

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

[3] R. Henderson, P. Miller, J. Stewart, A. Traxler, and R. Lindell, Item-level gender fairness in the force and motion conceptual evaluation and the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **14**, 020103 (2018).

[4] S. J. Pollock, Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA, AIP Conf. Proc. **1064**, 171 (2008).

[5] P. Eaton, K. Johnson, B. Frank, and S. Willoughby, Classical test theory and item response theory comparison of the brief electricity and magnetism assessment and the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **15**, 010102 (2019).

[6] A. S. Beavers, J. W. Lounsbury, J. K. Richards, S. W. Huck, G. J. Skolits, and S. L. Esquivel, Practical considerations for using exploratory factor analysis in educational research, Pract. Assess. Res. Eval. **18**, 1 (2013).

[7] PhysPort, Security FAQ for the Assessment Data Explorer, 2017 (accessed November 20, 2017), URL https://www.physport.org/DataExplorer/SecurityFAQ.cfm#2q3.

[8] M. W. McColgan, R. A. Finn, D. L. Broder, and G. E. Hassel, Assessing students' conceptual knowledge of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **13**, 020121 (2017).

[9] M. Tavakol and R. Dennick, Making sense of Cronbach's alpha, Int. J. Med. Educ. **2**, 53 (2011).

[10] K. D. Kubinger, On artificial results due to using factor analysis for dichotomous variables, Psychol. Sci. **45**, 106 (2003).

[11] T. F. Scott and D. Schumayer, Students' proficiency scores within multitrait item response theory, Phys. Rev. ST Phys. Educ. Res. **11**, 020134 (2015).

[12] L.-J. Weng and C.-P. Cheng, Parallel analysis with unidimensional binary data, Educ. Psychol. Meas. **65**, 697 (2005).

[13] R. B. Cattell, The scree test for the number of factors, Multivariate Behav. Res. **1**, 245 (1966).

[14] T. A. Brown, *Confirmatory Factor Analysis for Applied Research*, 2nd ed. (Guilford Publications, New York, NY, 2015).

[15] R. L. Gorsuch, *Factor Analysis* (L. Erlbaum Associates, Hillsdale, NJ, 1983).

[16] W. Revelle, psych: Procedures for Psychological, Psychometric, and Personality Research, Northwestern University, Evanston, Illinois, 2018, URL https://CRAN.R-project.org/package=psych, R package version 1.8.4.

[17] R. I. Jennrich and P. F. Sampson, Rotation for simple loadings, Psychometrika **31**, 313 (1966).

[18] H. F. Kaiser, An index of factorial simplicity, Psychome-trika **39**, 31 (1974).

[19] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using Multivariate Statistics* (Pearson, Boston, MA, 2007), Vol. 5.

[20] Y. Rosseel, lavaan: An R package for structural equation modeling, J. Stat. Softw. **48**, 1 (2012).

[21] P. Eaton and S. D. Willoughby, Confirmatory factor analysis applied to the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. **14**, 010124 (2018).