PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **19,** 010114 (2023)

Development and validation of a conceptual survey instrument to evaluate senior high school students' understanding of electrostatics

Shuaishuai Mi
School of Physics and Electronics, Shandong Normal University, Jinan, Shandong 250358, China

Jianqiang Ye
College of Chemistry and Materials Engineering, Wenzhou University, Zhejiang 325035, China

Yan Li
School of Education and Psychological Science, University of Jinan, Jinan, Shandong 250024, China

Hualin Bi[*]
College of Chemistry, Chemical Engineering and Materials Science,
Shandong Normal University, Jinan, Shandong 250014, China

(Received 26 January 2022; revised 27 June 2022; accepted 14 December 2022; published 28 February 2023)

This study developed and validated an instrument to investigate senior school students' understanding of electrostatics and provide a cognitive diagnostic assessment of their strengths and weaknesses on the related concepts (e.g., electric charge). The instrument included 20 four-tier multiple-choice items and the development process is organized around two activities: the development of the instrument and its validation. The development step defined the secondary concepts and designed the items using the misconceptions related to them. In the validation step, the instrument was applied to 1850 senior high school students from nine schools in two provinces in China, and the collected data were analyzed using the CDM package in R language. This step ensures that the diagnostic reports represent students' conceptual understanding reliably and validly by selecting the best model, analyzing item quality, overall test reliability, and the instrument's structure. The instrument can provide the percentage of students in the test population who possess certain combinations of concepts, the percentage of students in the test population possessing individual concepts, and the fine-grained size of concept proficiency information, which can be integrated as one completed report to issue to students, teachers, and parents to demonstrate students' status of conceptual understanding related to electrostatics. In addition, the construct induced from the diagnostic results can also be aggregated to the classroom, schools for instruction planning or low-stakes decision making, or infer a learning sequence.

DOI: 10.1103/PhysRevPhysEducRes.19.010114

I. INTRODUCTION

Electrostatics played an important role in transforming research scales in physics laboratories and presenting science to the public [1]. Learning about electrostatics can help people explain daily phenomena related to it, such as static cling, pollination, photocopying, lightning, and sparks [2].

As for students though electrostatics is only a small part of electromagnetism, its role in science education is important, as it is a prerequisite to studying electrokinetics [3,4]. In addition, understanding the concepts and principles of electrostatics is a prerequisite for conceptual reasoning in solving electrostatics problems [5,6]. Therefore, students must attain a conceptual understanding of electrostatics.

Studies have demonstrated that students have above-average difficulties with electrostatics due to its complexity and a high degree of abstraction [7,8], often lack knowledge regarding the meaning of terms related, or even have many misconceptions and learning difficulties on electrostatics topics (e.g., electric charge or electric field) [9,10]. Grasping these misconceptions and difficulties can describe students' understanding before and after instruction, providing relevant information on the effectiveness

---

[*]Corresponding author.
bihualin@sdnu.edu.cn

of teaching [11–13]. Furthermore, teachers can utilize the information about students' misconceptions of classroom events (preparing students to overcome these misconceptions), while curriculum developers can incorporate the information about students' misconceptions of teaching materials [14,15]. Therefore, to improve instruction on electrostatics in senior schools, it is necessary to develop an instrument to diagnose students' conceptual understanding.

In contrast to other physics topics, there have been few instruments for diagnosing students' conceptual understanding related to electrostatics [16]. A handful of instruments developed to assess student learning of electromagnetism have shortcomings in diagnosing students' performance on conceptual understanding due to the underlying theory and suitability. Therefore, this study intends to develop a conceptual test instrument with qualitative methods to diagnose students' conceptual understanding of electrostatics and perform statistical analyses on the quantitative student data to validate this test instrument.

II. BACKGROUND

A. Theoretical background: Electrostatics

Matter can have a physical property called electrical charge [17,18]. Electrostatics concerns static charges and the forces they exert on one another [19]. Coulomb's law enables the calculation of force that electric charge exerts [20,21]. A region where an electric charge experiences an electric force acting upon it is defined as an electric field [22]. The electric field is assumed to be the "agent" by which the force is manifested on charged particles, and it provides a convenient "mechanism" for dealing with electrical phenomena. Electric potential and electric potential energy use scalar quantities rather than through the vector nature of the electric field to describe the electric field, and this is more useful for many problems than force [17]. Capacitors play an important role in storing charge (and hence electrical energy) and could be used to do work or examine the electric field, and capacitance provides a "figure of merit" for how effective any type of capacitor is at holding charge [17].

B. Students' misconceptions

As mentioned above, students should be proficient in electric charge, electric field, Coulomb's law, electrical potential energy, electric potential, and capacitance to grasp electrostatics. Although few papers have been published on the topic, some researchers have investigated students' misconceptions (or difficulties) of the above concepts in electrostatics in detail. For example, students conceive that the force of attraction or repulsion of a small charged object exerted on a large one is smaller than that of a large one exerted on a small one when a small and a large charge are separated by a distance [19]. These misconceptions provide the basis for the design of incorrect (though tempting) distractors of items in the instrument [23], and Table I summarizes the misconceptions or difficulties of students when learning the concepts mentioned above.

C. Previously developed test instruments

Several instruments have been developed to measure students' conceptual understanding of electricity and

TABLE I. Students' misconceptions or difficulties when learning electrostatics concepts.

Concept	Misconceptions or difficulties	References	Item
1. Electric charge	1.1 No free charge, so no electrical current means no electrical field inside the insulators	Bilal and Erol [9] and Viennot and Rainson [24]	14
	1.2 Electrostatic objects cannot attract neutral objects	Hermita *et al.* [25]	10
	1.3 Students believe that there would be no electric field in the absence of electric charge	Taşkın and Yavaş [10]	6
	1.4 Charged objects have only one type of charge rather than consisting of an imbalance of opposite charges	Otero [26], Siegel and Lee [27]	10
	1.5 Students believe that charges do not transfer between conductors with charges of the same sign	Guruswamy *et al.* [16]	2
	1.6 Students think that there would be a transfer between oppositely charged conductors until one of the conductors became neutral	Guruswamy *et al.* [16]	3
	1.7 Students do not know electric charges have quantum properties	Yildiz [28]	14
	1.8 The charges on the two metal objects remain the same after touching regardless of the signs of the initial charges	Guruswamy *et al.* [16]	4
	1.9 A neutral object has no charge	Hermita *et al.* [25]	10
	1.10 Friction is the (only) cause of static electricity	Siegel and Lee [27]	1

(Table continued)

TABLE I. *(Continued)*

Concept	Misconceptions or difficulties	References	Item
2. Electric field	2.1 Electric field lines are found to be confused with the line of trajectory	Maloney *et al.* [29]	7
	2.2 Students perceived electric field lines as real entities; they believe that the electric field had volume, force, or density	Taşkın and Yavaş [30], Pocoví [31], Törnkvist *et al.* [10]	12
	2.3 An interpretation of formulae as if the quantities at the right of the equal sign were the cause of the quantities to the left ($E = F/q$)	Tembani [32]	9,11
	2.4 Students do not establish a clear difference between the concepts of field intensity and electric force	Taşkın and Yavaş [10], Furió and Guisasola [33], Chang [34]	3,18
	2.5 Charged particles in a uniform electric field move with a constant velocity	Bilal and Erol [9], Taşkın and Yavaş [10], Aidoo *et al.* [35]	7
	2.6 Any charged particle independent of the charge polarity moves in the direction of the electric field	Bilal and Erol [9], Taşkın and Yavaş [10], Aidoo *et al.* [35]	7
	2.7 Particles moving in the opposite direction of the field always slow down	Bilal and Erol [9], Taşkın and Yavaş [10], Aidoo *et al.* [35]	13
	2.8 Only charged particles on the electric field line are affected by electric fields (there is no force acting on the charge because there is no line passing through it)	Bilal and Erol [9], Taşkın and Yavaş [10], Aidoo *et al.* [35]	5,13,17
	2.9 There are no electric fields in the gaps between electric field lines	Taşkın and Yavaş [10]	5
	2.10 Students tend to believe that the electric field does not change when a new positive charge is added to the system, while the field would decrease when a negative charge is added	Viennot and Rainson [36]	11
	2.11 Students think electric field lines can cross each other and form sharp boundaries	Törnkvist *et al.* [31]	12
	2.12 Field lines can begin and end anywhere, and there are a finite number of field lines	Maloney *et al.* [29], Rainson *et al.* [37]	12
	2.13 Electric forces act at a distance with no necessary medium	Furió and Guisasola [33] and Galili [38]	20
	2.14 Students cannot discuss electric force and field superposition in terms of vector component additions	Furió and Guisasola [33], Cao and Brizuela [39]	3,6
	2.15 Students cannot predict the behavior of charged particles under the influence of an electric field	Galili [39]	7,13
	2.16 Students fail to reflect the density of field lines to the magnitude of electric force	Maloney *et al.* [29], Törnkvist *et al.* [31]	5
	2.17 Confusion about the vector/scalar properties of electric fields	Chang [34]	3
3. Coulomb's law	3.1 Students believe larger "objects" (in charge magnitude) exert larger forces than smaller "objects."	Bilal and Erol [9], Muthiraparampil [19], Maloney *et al.* [29],	4
	3.2 Confusion on both the effect of the magnitude of the charges and the distance of separation	Hermita *et al.* [25], Maloney *et al.* [29]	2,6,14
4. Electrical potential energy	4.1 The larger the path on the equipotential surface, the larger the work needed	Bilal and Erol [9], Maloney *et al.* [29], Lindsey [40]	8
	4.2 Students fail to identify and explain instances of positive, negative, and zero work	Doughty *et al.* [41]	8
	4.3 Students fail to identify work based on electric field line diagrams	Doughty *et al.* [41]	17
	4.4 Students fail to grasp the relationship between electric potential and energy	Chang [34]	16

(Table continued)

TABLE I. *(Continued)*

Concept	Misconceptions or difficulties	References	Item
5. Electric potential	5.1 Students cannot deduce the direction of the electric field from a change in potential	Maloney *et al.* [29] and Hammer [42]	9
	5.2 Associate relatively high and low potential to positively and negatively charged particles, respectively	Hazelton *et al.* [43]	16
	5.3 Most students cannot derive either energy differences or electric fields from the distribution of equipotential lines	Maloney *et al.* [29]	17,18
	5.4 Students confuse representations of equipotential lines and electric field lines	Chang [34]	13
	5.5 Students omit the idea of difference (Δ) which delays the adoption of the formula ($E = \Delta V/\Delta X \neq V/x$)	Chang [34]	9
	5.6 Confusion about the vector/scalar properties of electric potential	Chang [34]	16
6. Capacitance	6.1 The capacitance concept has no meaning for uncharged bodies	Guisasola *et al.* [44,45]	15
	6.2 Students are not aware of what goes on during the process of charging a body	Guisasola *et al.* [46]	19
	6.3 Inserting an insulator between two conductor parallel plates reduces the capacitance of the system due to preventing charge transfer from one plate to another, preventing the electrical current	Bilal and Erol [9]	15

magnetism, and some items in these instruments can diagnose students' conceptual understanding of electrostatics. The Brief Electricity and Magnetism Assessment (BEMA) developed by Chabay and Sherwood [47] was designed to assess student understanding of basic electricity and magnetism concepts (circuits, electrostatics, magnetic, and forces) covered in college-level calculus-based introductory physics courses. It is validated using classical test theory (CTT) by Ding *et al.* [48]. The Conceptual Survey of Electricity and Magnetism (CSEM) is developed to assess general physics (algebra and calculus-based physics) students' understanding of topics in electricity and magnetism (e.g., electrical content, particularly on static electricity [49]) [29]. Like BEMA and CSEM, the Colorado Upper-Division Electrostatics (CUE) assessment [50] and the Electricity and Magnetism Conceptual Assessment (EMCA) [51] are also developed to assess algebra and calculus-based physics students' understanding of topics in electricity and magnetism. However, the topics covered in these two assessments might differ slightly between BEMA and CSEM. Additionally, the CSEM, CUE, and EMCA are also validated using CTT.

Studies have demonstrated that all existing instruments are developed to assess undergraduate students' understanding of electrostatics rather than senior high school students. At the same time, electrostatics is only one of the physics topics covered in these instruments. Additionally, these instruments are validated using CTT. Finally, all these instruments provide a total score for each student, and then students' performances on conceptual understanding are measured through the pre-post score change or the normalized gain [52].

However, the topics and content depth of electrostatics courses at the senior high school level lack depth and breadth compared to the undergraduate level. For example, in senior high school, students will not learn about the "*electric field strength due to surface and volume charge distributions*" and "*Gauss' law*." Therefore, selecting items related to electrostatics in the above instruments to design an instrument directly might be inappropriate for senior high school students due to the topics and content depth. In addition, linking subcomponents from different instruments directly might lead to serious issues about reliability and validity [53].

CTT is "a measurement theory which consists of a set of assumptions about the relationships between actual or observed test scores and the factors that affect these scores, which are generally referred to as error" [54]. Under CTT, item difficulty and item discrimination indices are group dependent, meaning the values of these indices depend on the group of examinees in which they have been obtained [55]. Another shortcoming is that observed and true test scores are test dependent, and this means observed and true scores rise and fall with changes in test difficulty [56]. Given that, the instruments which CTT validates might suffer from these theoretical shortcomings.

Unlike CTT, item response theory (IRT) is a system of models that defines one way of establishing the correspondence between latent variables and their manifestations [57]. In IRT, persons and items are located on the same continuum [58] and thus can ensure the item difficulty and discrimination indices are independent of examinees. It can also obtain a set of personal ability estimates independently of the items used in an instrument [59]. Therefore, to address the shortcomings of CTT, some studies used IRT to validate the above instruments. For example, Ding [60] reanalyzed the construct of the Brief Electricity and Magnetism Assessment (BEMA) through the unidimensionality, invariance of the Rasch analysis (the simplest IRT model, see von Davier [61]) and found that several items may need further revision. In addition, the CSEM and BEMA were also reanalyzed using multidimensional item response theory [62,63].

In science education, students are expected to experience several processes of overcoming misconceptions related to specific concepts (e.g., work and electric fields; see Furió and Guisasola [33]). Then, they may form a scientific conception about a specific topic in physics (e.g., electrostatics). In other words, to establish a scientific conception of a specific topic in physics, students must grasp several secondary concepts related to this topic (e.g., work and electric fields vs electrostatics). Compared to an analysis of an overview of students' conceptual understanding related to a physics topic using a total score, teachers prefer to analyze conceptual learning related to a specific physics topic (e.g., electrostatics and circuits) by identifying proficiency (or grasping) states and constructs (links between concepts) of concepts related to this topic precisely [64–66]. Additionally, the analytic results can be used for classification-based decision making (e.g., to determine the best exercise strategy for the student, see McDermott and Shaffer [67]; for curriculum development, see McDermott and Shaffer [68]).

Though IRT can overcome some shortcomings of CTT to a certain extent, these models cannot provide more specific diagnostic and constructed information relevant to students' conceptual understanding of electrostatics by providing single overall scores [69]. Some researchers have used interviews, observations, and analyses of the protocols of learners' performance to obtain specific diagnostic information (difficulties or misconceptions) about students' conceptual understanding. However, these methods also suffer from shortcomings, such as the inability to investigate the construction between secondary concepts [70] and slow speed [71].

D. Cognitive diagnosis assessment

Cognitive diagnosis assessments (CDA) estimate the possession of attributes underlying a global ability using a class of discrete latent variable models [72,73]. In this context, the global ability can be viewed as the mathematical abilities of students of a given age; then, a CDA can be used to estimate proficiency in basic mathematical operations (such as *converting a whole number to a fraction* or *separating a whole number from a fraction*) of students underlying their mathematical abilities [72,74]. Global ability can also be viewed as the science competency proficiency of students of a given age, and a CDA can provide information about the possession of diagnostic attributes (such as *basic science knowledge*, *using models*, or *reasoning*), which are parts of science competency proficiency [75]. Based on the literature, if proficiency in a global ability (mathematical abilities or science competency proficiency) requires students to become more proficient in basic elements (such as *converting a whole number to a fraction*, *separating a whole number from a fraction*, *basic science knowledge*, *using models*, or *reasoning*) of this global ability, then a CDA can provide information about the proficiency in these fundamental elements. However, nothing is inherent in formulating a CDA that prevents the meaning of the general ability from being broadened to include other constructs, such as a specific topic in physics [76]. Therefore, when the general ability is taken as a given topic in physics (e.g., electrostatics) [77,78], a CDA can provide information on which secondary concepts have been grasped [69]. Furthermore, a CDA can provide the secondary concepts' correlations and hierarchies [79], and recently some researchers have also used it to monitor student's progress toward learning goals [80].

Currently, there is a lack of instruments to diagnose students' conceptual understanding of electrostatics built from a CDA [69]. Some researchers attempted to fit a model for cognitive diagnosis to test data from assessments originally designed under a different measurement (e.g., CTT or IRT) framework, and this process is termed "*retrofitting*" [79]. For example, Mirzaei *et al.* retrofitted the reading section of the International English Language Testing System (IELTS) with CDAs [81]. Because of the original intended use of these instruments (e.g., BEAM and CUE), retrofitting cannot be expected to provide sufficient diagnostic information, at least not in its entirety. The glaring disparity between the theories used in designing and analyzing an instrument casts doubts on the diagnostic value of cognitively based analysis of CTT-based or IRT-based data [69].

Therefore, this study aims to determine the extent to which the diagnostic information estimated by a CDA provides a discriminant, accurate, and reliable method for determining senior high school students' conceptual understanding or whether the developed instrument based on CDAs is valid for use in diagnosing students' conceptual learning of electrostatics.

III. INSTRUMENT DEVELOPMENT, CONTENT, AND FORMAT

A. Instrument development

The instrument development usually involves several steps, starting with the educational and didactic experts defining the secondary concepts (e.g., work and electric fields) related to electrostatics [82]. As mentioned in Sec. II A, if students have strengths in electric charge, electric field, Coulomb's law, electrical potential energy, electric potential, and capacitance, then they might be proficient in electrostatics.

The performance of students becoming aware of scientific conceptions is to overcome misconceptions [83]; therefore, if a student can identify and abandon tempting distractors arising from misconceptions about an item, they might grasp concepts. As demonstrated in Table I, students might hold several misconceptions about a concept (e.g., electric charge and electric field); in addition, the relations among concepts in electrostatics are also not independent but closely related [34]. Such circumstances allow us to simultaneously assess more than one concept or misconception in a single item, and in CDAs, these relations are stored in a Q matrix.

A Q matrix is an item-by-concept matrix that specifies the concepts required to answer each item correctly in the instrument [72,84]. Ideally, the Q matrix presents conditions that guarantee the identifiability or local identifiability of the conjunctive CDAs, such as having items requiring only a single concept, concepts measured by at least two or three items, and two items with identical concept requirements for every concept defined in the Q matrix [85].

The second step in the instrument development includes preparing a Q matrix and item design, which are two simultaneous processes. The detailed process is described as follows:

1. We collaborated with a focus group consisting of a high school physics teacher and a science education researcher to develop questions for the instrument. The relationships between concepts are defined based on the physics equation in textbooks [86]. For example, the magnitude of the electric field at a point is defined using the ratio of a charge q experiencing a force F at this point to the q, and Coulomb's law enables the calculation of forces; therefore, the groups thought there ought to be relationships between either pair of these three concepts. Then the focus group distributes the misconceptions in Table I to the relationships defined in the previous step. For example, there is a relation between *electric charge* and *Coulomb's law*, and the misconceptions about these two concepts are assigned to the relations between these two concepts. Third, the focus group designs an item that assesses the misconceptions about electric charge and Coulomb's law. Considering the number of misconceptions (more than ten about electric fields), one relation might be assessed by more than one item, though these items might assess different misconceptions. In addition, a concept must be measured by at least two items [85]. In addition, the instrument length might affect students' cooperation rates, and the longer the length, the fewer respondents started and completed the questionnaire [87]. Therefore, considering the above two aspects, to minimize the instrument length, in the fourth step, the focus group assesses as many relations and misconceptions in an item as possible.
2. Once the items have been designed, the focus group then designs distractors arising from misconceptions about each item. Generally speaking, the focus group assumes that if a misconception is held by students, what kind of answer would students give to an item through their response process based on the misconception. Then this answer will be set as a distractor for this item (see example item in Sec. III C). To reduce the possibility of students selecting a correct choice based on guessing, the focus group provides as many distractors as possible for an item (in some items, there are more than ten choices).
3. After developing the test items, the test items and Table I were submitted to a high school physics teacher (who did not belong to the focus group) who marked whether the test items required students to overcome the misconceptions that the focus group thought. The consistency coefficient between these two tables is calculated, which was 0.85. Then the focus group and physics teacher discussed the inconsistencies between the two tables until they agreed on whether the misconceptions were examined in a test item.
4. If an item assesses a misconception (e.g., misconception 1.1, 1.2 in Table I) about a concept (e.g., electric charge) in electrostatics, then this item is thought to assess this concept. The focus group specified which concepts were required to obtain a positive response in each test item, using a Q matrix according to the item designed, and the information was provided in the fourth column of Table I [72]. For example, the sample item (see Sec. III C) involves misconceptions 1.8 and 3.1, which means that this item examined electric charge and Coulomb's law. The row and column (corresponding to this item and these two concepts) are marked by 1 (see Table II, row 5, column 1; row 5, column 3). The focus group assumed students' response process to the items, and students might provide an incorrect (or correct) choice using misconceptions (or correct concepts) that are inconsistent with the focus group's assumed choice. Therefore, though there is a consensus between the focus group and the physics teacher, the Q matrix might not represent the students' real response to the items. Therefore, the Q matrix will be validated by the reasons for the selection of a choice provided by the students

TABLE II. The Q matrix for the electrostatics test.

Item	Electric charge	Electric field	Coulomb's law	Electrical potential energy	Electric potential	Capacitance
1	1	1	1	0	0	0
2	1	0	1	0	0	0
3	1	1	1	0	0	0
4	1	0	1	0	0	0
5	0	1	1	0	0	0
6	1	1	1	0	0	0
7	0	1	0	0	1	0
8	0	1	0	1	1	0
9	0	1	0	0	1	0
10	1	1	0	0	0	0
11	1	1	0	0	0	0
12	1	1	0	0	0	0
13	0	1	1	0	1	0
14	1	0	1	0	0	0
15	0	0	0	0	0	1
16	0	1	0	1	1	0
17	0	1	0	1	1	0
18	0	1	0	0	1	0
19	0	0	0	0	0	1
20	0	1	0	0	0	0

(for the validated process, see Sec. V for details). Table II presents the final Q matrix after validation.

5. Theoretically, the instrument consisting of items can be delivered to students to diagnose their conceptual understanding of electrostatics using students' response data. The diagnostic results outputted by CDAs might not be consistent with students' conceptual understanding (e.g., correct answers due to correct reasoning or due to incorrect reasoning, see Caleon and Subramaniam [88]); therefore, the focus group introduces a reason tier (R tier), which includes a multiple-choice set of reasons for the answer in the items, to each item [89,90]. Additionally, students' response data to the R tier can be used to validate the instrument in this study.
6. To avoid the omission of misconceptions, in the multiple-choice content questions and its R tier, the focus group adds two fill-in-the-blanks choices that ask students for an answer in case all answers provided cannot represent the students' answers and provides a reason for the answer in case all reasons provided in R tier cannot represent the students' reasons [91].
7. The R tier can measure the reason behind students' choices, which can be used to validate the instrument, but it cannot determine whether students' incorrect answers arise from a lack of knowledge of electrostatics [92]. Additionally, the students who lack knowledge may be inappropriate for the instrument. In this study, to identify such students, the confidence ratings for the answers to the items and R tier are introduced as additional tiers [88].

B. Content

The instrument developed in this study is a 20-question conceptual test that assesses six concepts with 42 misconceptions in Table I. A brief description of all the questions is given in Table III along with concepts related to that question. One sample question from the instrument is demonstrated in Sec. III C, and the full instrument is given in the Appendix.

As mentioned above, if an item involves a misconception about a concept in electrostatics, then this item is thought to assess this concept. Table III illustrates that the misconceptions involved in items do not correspond exactly to the concepts assessed in an item. For example, item 5 involves misconceptions 2.8, 2.9, and 2.16 (see Table I), which means this item assesses only the electric field, but Table III demonstrates that this item also assesses Coulomb's law. In this study, the focus group assumes the misconceptions involved in an item, meaning that students might respond to the item using a different process with different misconceptions. The concepts assessed in an item illustrated in Table III represent the concepts assessed by items exactly after validation using reasons for the selection of a choice provided by the students (see Sec. V), and this agrees with the Q matrix in Table II.

C. Format

Twenty multiple-choice questions are designed by focus groups in the previous subsection (see Sec. III A), and the questions in the instrument are organized in the form of a four-tier multiple-choice (4TMC) test, which includes an

TABLE III. The brief description of all questions along with concepts related to that question.

Q no.	Description	Assess concepts	Involve misconceptions (see Table I)
1	Charging a conductor by induction	Electric charge, electric field, Coulomb's law	1.10
2	The calculation of the magnitude of electric force after two same signs of charges touching	Electric charge, Coulomb's law	1.5, 3.2
3	The calculation of the magnitude of electric force after two opposite signs of charges touching	Electric charge, electric field, Coulomb's law	1.6, 2.4, 2.14, 2.17
4	Keep the distance between two same signs of charge, then calculate the magnitude of electric force after one of them touches with another same sign of charge	Electric charge, Coulomb's law	1.8, 3.1
5	The relationship between the magnitude of electric force with electric field lines	Electric field, Coulomb's law	2.8, 2.9, 2.16
6	Electric field at a point due to several point charges	Electric charge, electric field, Coulomb's law	1.3, 2.14, 3.2
7	Direction of motion of a point charge in two uniform electric field	Electric field, electric potential	2.1, 2.5, 2.6, 2.15
8	The work done of electric force in an electric field generated by two-point charges with the same signs	Electric field, electric potential energy, electric potential	4.1, 4.2
9	Calculate the magnitude of the electric field from the equipotential lines	Electric field, electric potential	2.3, 5.1, 5.5
10	Predict the motion due to the induction	Electric charge, electric field	1.2. 1.4, 1.9
11	Test charges effects on electric field	Electric charge, electric field	2.3, 2.10
12	The visual representation of the electric field using electric field lines	Electric charge, electric field	2.2, 2.11, 2.12
13	Predict the motion of a point charge in the field which is generated by a positive charge	Electric field, Coulomb's law, electric potential	2.7, 2.8, 2.15, 5.4
14	The calculation of the magnitude of the electric field generated by a point charge	Electric charge, Coulomb's law	1.1, 1.7, 3.2
15	The influencing factors of the capacitance of a capacitor	Capacitance	6.1, 6.3
16	The relationship between electric potential, electric potential energy, and the electric field generated by two point charges with opposite signs	Electric field, electric potential energy, electric potential	4.4, 5.2, 5.6
17	The work done of electric force, and the change of electric potential energy with the motion of a charge, in an electric field generated by a negative point charge	Electric field, electric potential energy, electric potential	2.8, 4.3, 5.3
18	To induce the magnitude of the electric field and the electric potential energy, from the equipotential lines	Electric field, electric potential	2.4, 5.3
19	The change of voltage at the terminals of the capacitor, and the change of the current flowing through the capacitor	Capacitance	6.2
20	the long-range nature of electric interactions.	Electric field	2.13

answer tier (multiple-choice questions, termed as the *A* tier), *R* tiers with additional tiers requiring students to specify confidence ratings separately for their choice of answers in the *A* and *R* tiers [88].

We illustrate a sample item (item 4, the diagram for it, see Fig. 1) demonstrating that the *A* tier lists all the possible choices based on misconceptions 1.8 and 3.1 (see Table I), and students can select one choice, or provide an answer in the fill-in-the-blanks choice when they think there is no correct choice from the existing options. The *R* tier lists the possible reasons for the students' selections in the *A* tier, and the blank space is provided for the students to write their reason in case they cannot find a suitable explanation for their answer among the choices in the *R* tier. Two six-point confidence scales (based on McKelvie [93]) were added below the *A* tier and *R* tier of each item in the instrument, with "1" and "6" corresponding to "Just Guessing" and "Absolutely Confident," respectively. All

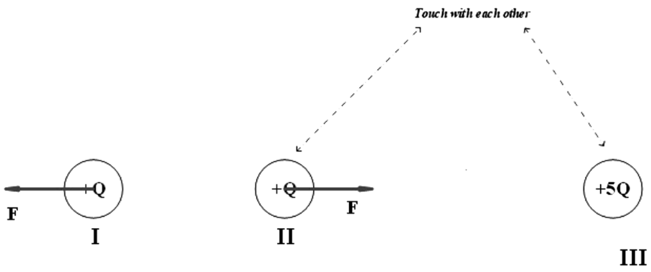

FIG. 1. Figure for Q4.

questions (question 1, the diagram for it see Fig. 5; question 2, the diagram for it see Fig. 6; question 3, the diagram for it see Fig. 7; question 4, the diagram for it see Fig. 8; question 5, the diagram for it see Fig. 9; question 6, the diagram for it see Fig. 10; question 7, the diagram for it see Fig. 11; question 8, the diagram for it see Fig. 12; question 9, the diagram for it see Fig. 13; question 10, the diagram for it see Fig. 14; question 12, the diagram for it see Fig. 15; question 13, the diagram for it see Fig. 16; question 14, the diagram for it see Fig. 17; question 15, the diagram for it see Fig. 18; question 16, the diagram for it see Fig. 19; question 17, the diagram for it see Fig. 20; question 18, the diagram for it see Fig. 21; question 19, the diagram for it see Fig. 22; question 20, the diagram for it see Fig. 23) of the instrument are given in the Appendix.

Example Question 4

Two small metal objects (I and II), each with a net charge of $+Q$*, exert a force of magnitude* F *on each other. The right-hand object (II) is now made to touch another metal object whose net charge is* $+5Q$ *and then brought it (II) back to the same distance apart from I. Then, what is the magnitude of the force on the metal objects I and II (the III is withdrawn)?*

(A) *F, F*
(B) *3F, F*
(C) *F, 3F*
(D) *3F, 3F*
(E) *Your answer:* ______

Q4.A Confidence Rating for Q4

(A) *Just Guessing*
(B) *Very Unconfident*
(C) *Unconfident*
(D) *Confident*
(E) *Very Confident*
(F) *Absolutely Confident*

Q4.B Scientific reasons for my answer for Q4

(A) *The charges on the two metal objects remain the same after touching; therefore, the I and II have* $+Q$ *charge, and the magnitude of the force exerted on each other are also equal to F.*
(B) *Larger "objects" (in charge magnitude) exert larger forces than smaller "objects," therefore, II exerts 3F on I, and I exert F on II.*
(C) *Your reason:* _______

Q4.C Confidence Rating for Q4.B

(A) *Just Guessing*
(B) *Very Unconfident*
(C) *Unconfident*
(D) *Confident*
(E) *Very Confident*
(F) *Absolutely Confident*

As students respond to the items, they might first select a choice from four options in *A* tier, and their answer might be signed as "0" or "1" for incorrect or correct responses, respectively. Then the reasons for the answer in the *A* tier might be selected or provided by students to validate the instrument. Finally, the *A*-tier and *R*-tier confidence ratings might also be provided to distinguish students who are not appropriate for the instrument, and these students might be removed as participants.

The instrument is typically given to students after learning about electrostatics and requires approximately 90 min.

IV. METHODS

A. Sample

We selected the participating students using convenience sampling. The students were from nine senior high schools in Shandong province, Beijing, and 2563 students participated in this study.

Some students specified the confidence ratings of all items as *just guessing* in the *A* or *R* tiers. The data of these students were regarded as invalid data and were removed. Finally, the data from 1850 eleventh graders remained. These students were aged from 16 to 18 ($M_{\mathrm{age}} = 16.86, \mathrm{SD}_{\mathrm{age}} = 0.65$), 65% were boys, and 35% were girls.

B. Test administration

Before the instrument was delivered to the students, they all had completed sessions related to electrostatics while following the physics textbooks. Additionally, before the test, they were told it was a diagnostic test, not an achievement test. Further, they were informed that the test results would not affect their school grades but would be used by their teachers in planning their remedial lessons. This reassurance is essential, so students do not provide socially desirable responses for the confidence ratings.

Then students read a guidance material developed to tell students some information about the instrument. In this material, students will understand the number of items in this instrument, the number of correct answers for each item (only one), the answer behavior allowed in each tier of items, and the maximum answer time (90 min). In this material, students will naturally know that they can provide additional answers in the *A* tier, and reasons in the *R* tier, if they think there is no correct answer or no reasons that can represent their reason. After students read the above material, we tested the students using the online data collection tool Wenjuanxing (see [94]). The test lasted approximately 90 min.

C. Cognitive diagnosis assessment models

CDA can provide information on which concepts (e.g., electric charge, electric field) have been grasped. To provide diagnostic information, statistical techniques, often referred to as CDA models or cognitive diagnostic models (CDM) in the measurement literature, have been devised [73,95–99].

To illustrate the difference between the CDA models, the notations used in the models are defined. Let X_{ij} be the response of student i to item j ($i = 1, \ldots, I$, $j = 1, \ldots, J$, $I = 1885$, $J = 20$ in this study), where 1 (or 0) on the ith row and jth column, denotes student i response item j correctly (or incorrectly). Additionally, let $\alpha_i = \{\alpha_{ik}\}$ be the students' binary concept vector ($k = 1, \ldots, K, K = 6$ in this study), where a 1 on the kth element denotes proficiency of concept k and 0 denotes nonproficiency of the concept. In addition, CDA models also require a cognitive design matrix that explicitly identifies each item's cognitive specification, termed as Q matrix (see Sec. III A).

In the DINA (deterministic inputs, noisy "and" gate) model, a student's concept vector and the Q matrix produce a latent response vector $\eta_i = \{\eta_{ij}\}$, where $\eta_{ij} = \sum_{k=1}^{K} \alpha_{ik}^{q_{jk}}$. The equation assumes a value of 1 if student i possesses all the skills required for item j and a value of 0 if the student lacks at least one of the required concepts. The "and" gate component of the DINA model refers to the conjunctive process in determining η_{ij} in that a correct response to an item requires the presence of all the prescribed concepts for the item [73].

Compared to the DINA model, in the DINO (deterministic input; noisy "or" gate) model, students should be proficient in at least one of the prescribed concepts for the item; from which they might provide a correct response to an item [99].

A model requiring that each concept is present to produce a correct response is termed a noncompensatory model, while a model allowing a deficit in one concept can be compensated for by a surplus in another concept is termed a compensatory model [100]. Therefore, the DINA is a noncompensatory model, while the DINO is a compensatory model.

If students' response process is completely deterministic (i.e., error free or nonstochastic), the latent response vector is identical to the manifest or observed response vector. However, because the underlying process is inherently stochastic, the latent response vector represents only an ideal response pattern. The DINA and DINO models thought that the noise in the process is due to slips and guessing parameters; that is, students who possess all the required concepts for an item can slip and miss the item, and students who lack at least one of the required concepts can guess and still answer the item correctly with typically nonzero probabilities [73,99].

The HO-DINA (higher order DINA) model links a latent (continuous) trait with several latent (binary) concepts in a hierarchical framework using the DINA model so that both types of information are available simultaneously [96,101]. Compared to the DINA model, this model also assumes that a higher level of trait produces the proficiency status of concepts.

In the rRUM model, the probability of a student i responding correctly to an item j is decided by the probability of a correct response for individuals with all requisite attributes (π_j^*), and a penalty parameter for missing a required concept $k(r_{jk}^*)$ [98,102]. Additionally, the π_j^* and the ($r_j k^*$) also can be calculated by slip and guessing parameters [103]. The rRUM is a refinement of the DINA model where the required concepts that are lacking influence the probability of a correct response [104].

Given the numerous CDA models that have been devised, De La Torre [97] proposed a framework for relating several CDA models, the G-DINA (generalized deterministic inputs, noisy "and" gate) model. In this model, the probability of student i with the concepts vector α_i answering j correctly can be decomposed into the sum of the effects due to the presence of specific concepts and their interactions (see details in De La Torre [97]), and the other CDA models can be obtained through setting some parameters to zero. For example, by setting all the interaction effects to zero, we obtain a model that will be denoted as the additive CDM (A-CDM). In the A-CDM model, becoming proficient in any concept increases the probability of success on an item, and its contribution is independent of the contribution of the other concept [97].

In summary, there are differences in the characteristics of CDA models. For example, the interaction of the concepts (noncompensatory or compensatory) [100], the different attributions about students' responses correctly [73,98], and the constructs between concepts [96]. It is difficult to select an appropriate CDA model with no information about the relationships between concepts related to electrostatics, but specific CDA models have more straightforward interpretations, are more stable, and can provide more accurate diagnostics information when used correctly [105].

In general, the most appropriate CDA model is selected from several alternative CDA models by evaluating the fit parameters [79,106]. Therefore, in this study, the best model selected from the following six options based on the input data, which includes the Q matrix (see Table II) and students' response data to each item (especially A tier, see Sec. III A) [107]: the DINA model [73]; the DINO model [99]; the A-CDM model [97]; the rRUM model [98,102]; the HO-DINA model [96]; and the G-DINA model [97].

V. VALIDATION

As mentioned above, selecting appropriate CDA models is the prerequisite to providing more accurate diagnostics information. Therefore, the first step to validate the

TABLE IV. Absolute fit and relative fit indices.

	Absolute fit		Relative fit		
Model name	SRMSR	RMSEA	AIC	BIC	LL
G-DINA	0.039	0.038	38923.127	39607.971	−19337.563
DINA	0.082	0.051	41447.964	41790.386	−20661.982
DINO	0.082	0.043	41475.607	41818.029	−20675.803
A-CDM	0.048	0.058	39400.123	39875.096	−19614.062
rRUM	0.056	0.068	39858.629	40333.602	−19843.315
HO-DINA	0.053	0.067	39661.119	40290.734	−19716.559

instrument is to select an appropriate CDA model based on model fit, item fit, and person fit.

Once the CDA model is selected, the instrument is validated from three aspects, respectively, students' overall understanding of all concepts identified, students' understanding of specific concepts, and their propensity for misconceptions [108]. The analytic process adapts the analytic framework provided by Jorion *et al.* [108] to the instrument.

At first, the item quality and test properties are investigated, which can identify the problematic items and measurements. Because CDA models are designed for diagnostic purposes (not specifically designed for measuring students' success rates in a test), they use different measures for defining an item as good or bad [109]. Additionally, CDA models examine item quality by determining item discrimination [110]. In addition, the reliability measures for the estimates obtained from the CDA models mostly agree on the estimated and true skill (termed classification accuracy) or between estimated skills from parallel assessments (termed classification consistency) [111]. Therefore, their measures are taken to indicate the reliability of the instrument [95,112]. This study also analyzed the item difficulty, discrimination, and test reliability from CTT and IRT to illustrate the item and test properties clearly [79,113].

Next, the structure of the instrument is evaluated. In CDA models, the relationships between items and concepts are specified in the Q matrix; therefore, the Q matrix is calibrated carefully based on the reasons students provided in R tier.

Last but not least, students' degree of proficiency in individual concepts is identified, including students' concept profiles and ability estimates [114]. These diagnostic output results are also confirmed by the reasons students provided in the R tier.

In this study, all the statistical analyses are conducted using the CDM package [115] in R language.

A. Selection of an appropriate cognitive diagnosis assessment model

Model fit, also termed test-level fit, is used to analyze whether the selected model fits the data entirely [116]. It includes absolute and relative model fits. Absolute fit evaluations determine whether the model at hand adequately fits the data [117], and relative fit evaluation selects the best-fitting model among a set of competing models when more than one model can adequately fit the data [117].

Regarding the absolute fit evaluation, the standardized root mean squared residual (SRMSR) [118] and the root mean square error of approximation (RMSEA) [119] were used. As for the relative fit evaluation, the −2 log-likelihood (LL), Akaike's information criterion (AIC), and Bayesian information criterion (BIC) [117,118] were used.

The item-level appropriate measures indicate whether the model fits individual items carried up [116]. Item-level fit measures include the RMSEA, which compares observed and predicted responses for different latent classes [120].

The person fit measures refer to the correspondence between an examinee's observed response pattern and their expected response pattern, given that their estimated skill profile is finally carried [79]. For the person fit, measures include the likelihood ratio test, which can identify aberrance from normal behavior in cognitive diagnosis models (one can assume the model used should be reconsidered if person fit analyses result in too many rejections; see [121]).

1. Results of absolute fit and relative fit evaluation

The SRMSR and the RMSEA of the squared residual indices were estimated to determine absolute fit (see Table IV). Optimal fit is reached when RMSEA and/or SRMSR are below 0.05 [114,122,123].

As illustrated in Table IV, the G-DINA, A-CDM, and DINO models have the best absolute fit. The G-DINA model has the best fit in terms of SRMSR and RMSEA among the models.

The CDA models were compared using the fit metrics of LL, AIC, and BIC. The results for the model relative fit indices are presented in Table IV. For each of these three statistics, the fitted model with the smallest value was selected among the competing models [117]. In terms of the LL, AIC, and BIC, the best fit was achieved by the G-DINA model, followed by the A-CDM and DINO models.

2. Results of item fit and person fit

Item fit was determined using the item level RMSEA index. In Table V, the G-DINA model tended to have the best item-level fit, with 15 items with RMSEA values below 0.05. The DINO and rRUM models have 13 items

TABLE V. RMSEA item fit indices.

Item	G-DINA	DINA	DINO	A-CDM	rRUM	HO-DINA
1	0.018	0.035	0.027	0.038	0.034	0.029
2	0.021	0.031	0.015	0.116	0.060	0.082
3	0.024	0.061	0.024	0.054	0.029	0.041
4	0.043	0.072	0.044	0.069	0.056	0.108
5	0.109	0.013	0.059	0.034	0.047	0.036
6	0.023	0.023	0.026	0.040	0.023	0.041
7	0.065	0.066	0.055	0.061	0.035	0.054
8	0.019	0.067	0.067	0.019	0.027	0.030
9	0.014	0.062	0.054	0.069	0.049	0.078
10	0.039	0.063	0.033	0.086	0.049	0.057
11	0.063	0.079	0.094	0.119	0.163	0.163
12	0.018	0.061	0.023	0.052	0.267	0.058
13	0.035	0.053	0.020	0.033	0.044	0.022
14	0.065	0.093	0.105	0.049	0.149	0.114
15	0.042	0.016	0.013	0.014	0.019	0.069
16	0.015	0.041	0.028	0.012	0.023	0.064
17	0.004	0.057	0.036	0.103	0.051	0.044
18	0.028	0.059	0.045	0.059	0.162	0.080
19	0.032	0.024	0.031	0.040	0.029	0.076
20	0.089	0.046	0.064	0.102	0.045	0.093

below the fit threshold. The A-CDM and DINA models had items below the fit threshold, respectively, nine and eight. The HO-DINA model had the worst item-level fit of all the models, with seven items below the fit threshold.

Person fit was assessed using the likelihood ratio test for aberrant behavior [121]. Table VI summarizes the proportion of students with misfits due to spuriously high and low scores, given their estimated skill profile. The G-DINA presented the lowest proportion of students with a misfit, whereas the DINA and DINO models had the highest.

The evidence from the absolute and relative fit indices favored the G-DINA model, followed by the A-CDM and DINO models. The item fit favored the G-DINA model, which had the most items with RMSEA values below 0.05, and the person fit results confirmed a better fit to the data for the G-DINA model than the other models. To emphasize the most parsimonious CDA model with a good fit to the data to provide more accurate diagnostics information, the results presented in the following subsection focus only on item and examinee estimates produced by the G-DINA model.

TABLE VI. Proportion of examinees with aberrant scores.

Model	Proportion of examinees with aberrant high scores	Proportion of examinees with aberrant low scores
G-DINA	0.08	0.12
DINA	0.14	0.14
DINO	0.14	0.14
A-CDM	0.12	0.13
rRUM	0.12	0.14
HO-DINA	0.12	0.14

B. Overall test and individual item properties

1. The results of item quality

The mean observed score on the instrument was 8.47 out of 20 ($\mathrm{SD} = 4.9$) or 42.4% correct. Table VII illustrates the item difficulty, item discrimination indices from the CDA model, CTT, and IRT.

With respect to the item discrimination index (IDI) from the CDA model, Table VII presents the IDI of each item estimated by the G-DINA model. The values of the IDI vary between 0.29 and 1, with an average of 0.81. As Robert Ebel [124] suggested, in terms of IDI, 0.40 and greater is very good, 0.30 to 0.39 is reasonably good, 0.20 to 0.29 is marginal, and below 0.19 is considered poor. Therefore, the IDI of the items has acceptable values except item 7, which means all items are accepted.

As for the item difficulties from CTT, the value ranged from 0.20 to 0.60, except for item Q7, which had a difficulty of 0.19. A lower difficulty value indicates a harder item; the values for all items but Q7 fell within the generally accepted range of item difficulties of 0.20 to 0.80. The range of item difficulties suggests that the instrument has a mix of easy, medium, and hard items that would be appropriate for a range of students who vary in proficiency regarding the knowledge domain of electrostatics. The item discrimination measure ranged from 0.28 to 0.64. Q7 and Q12 were again an exception, with item discrimination values of 0.11 and 0.05, which was outside the recommended range [125]. Apart from Q7 and Q12, the item discrimination values constitute reasonable evidence that each item's score is positively related to the overall proficiency represented by total performance on the instrument.

As for IRT, the reason for selecting the 2PL model is that the guessing parameter (i.e., the 3PL model) did not improve the model fit, and this result suggests that guessing was not a necessary modeling addition for the instrument, which means that it might be difficult for the students to choose the correct choice just based on guessing. The item is said to be good if the difficulty index is more than −2.00 and less than 2.00 [126]. Table VII illustrates that the difficulty index of items is in the range $-0.56 < difficulty < 1.46$, which means that the difficulty level of the developed items has met the difficulty index value category, so it can be said that all the items are good. As for item discrimination, the measure ranged from 0.60 to 2.34. Q7, Q11, and Q12 were an exception, with item discrimination values of 0.28, 0.06, and 0.60, respectively, which was outside the recommended range (less than 0.64 indicates low discrimination, see Baker [127]). Apart from Q7, Q11, and Q12, the item discrimination values constitute reasonable evidence that an item can differentiate between students having abilities

TABLE VII. The item quality indices from the CDA model, CTT, and IRT. Note that an analysis of students' response data to the instrument was performed using the one- (1PL), two- (2PL), and three-parameter logistic (3PL) models. The model-data fits of the three models were compared using the Akaike information criterion (AIC) fit statistic; the 2PL model had the best fit (1PL, AIC = 42297.64; 2PL, AIC = 41308.7, 3PL, AIC = 41310.4).

Item	Item discrimination (CDA model)	Item difficulty (CTT)	Item discrimination (CTT)	Item difficulty (2PL)	Item discrimination (2PL)
Q1	0.96	0.54	0.41	−0.26	1.12
Q2	0.71	0.53	0.52	−0.31	1.76
Q3	0.93	0.33	0.48	0.89	1.43
Q4	0.89	0.39	0.52	0.57	1.63
Q5	0.59	0.55	0.48	−0.33	1.48
Q6	1.00	0.46	0.55	0.15	1.78
Q7	0.29	0.19	0.11	1.46	0.28
Q8	1.00	0.29	0.59	1.36	2.02
Q9	0.82	0.40	0.35	0.48	0.96
Q10	0.95	0.43	0.39	0.33	1.11
Q11	0.73	0.45	0.28	0.20	0.60
Q12	0.92	0.20	0.05	1.41	0.06
Q13	1.00	0.50	0.55	−0.15	1.90
Q14	0.92	0.36	0.64	0.81	2.34
Q15	0.59	0.59	0.45	−0.56	1.47
Q16	0.96	0.31	0.56	1.19	1.96
Q17	1.00	0.52	0.51	−0.22	1.64
Q18	0.93	0.24	0.36	1.30	0.89
Q19	0.58	0.59	0.47	−0.57	1.51
Q20	0.48	0.60	0.40	−0.54	1.17

TABLE VIII. Reliability at test and attribute levels.

	Pattern level	Electric charge	Electric field	Coulomb's law	Electrical potential energy	Electric potential	Capacitance
Classification accuracy	0.85	0.91	0.99	0.99	0.93	0.91	0.96
Classification consistency	0.78	0.85	0.99	0.98	0.90	0.86	0.94

below the item location and those having abilities above the item location.

As noted above, item Q7 was identified as a problematic item by the CDA models, CTT, and IRT, and Q11 and Q12 were identified by CTT or IRT. As the purpose of the instrument was to determine the strengths and weaknesses of proficiency in concepts related to electrostatics and did not aim to measure students' academic achievement, no high- or low-level outliers were deleted at the item difficulty and discrimination from CTT and IRT [109]. Therefore, Q12 and Q11 are preserved. Item Q7 is also preserved after reviewing the reasons students provided to answer this question (see Sec. VI).

2. *Reliability of test*

Table VIII summarizes the agreement (classification accuracy and classification consistency) at the attribute pattern level and the attribute level [111]. Table VIII illustrates that the estimated pattern-level accuracy and consistency measures were 0.85 and 0.78, respectively. All the attribute-level accuracy measures were larger than 0.90, and the consistency measures were larger than 0.85.

As Johnson and Sinharay [128] suggest, in terms of agreement, 0.90 and greater refers to excellent reliability, 0.80–0.90 refers to very good reliability, 0.65–0.80 refers to good reliability, 0.50–0.65 refers to fair reliability, 0.25–0.50 refers to poor reliability, and < 0.25 refers to no reliability. This suggests very good or good reliability of the posterior modes for this assessment. This means that the instrument developed in this study can provide accurate skill and reliable proficiency classification, which means it can distinguish proficient from nonproficient students.

CTT provides Cronbach's alpha as an index of the total score reliability, where reliability means that a student's total score would be nearly the same if one could administer the test multiple times to the same student. The alpha value for the instrument was 0.85, which indicates good reliability for an assessment used for low-stakes purposes [108].

The EAP reliability index (based on expected *a posteriori* parameter estimates) from IRT can be viewed as the amount by which the measurement process has reduced the uncertainty in the latent ability of the student [129], and the value of it was 0.85, which can be regarded as fairly satisfying.

All reliability indices provided by CDA models, CTT, and IRT indicate that the instrument developed in this study might provide reliable results about students' conceptual understanding of electrostatics.

C. Evaluation of the structure of the instrument

In this study, the experts modeled the physics tasks as functions of proficiency and nonproficiency of one or more concepts [130], and the relationships between tasks and concepts are specified in the Q matrix (see Table II). Students' response processes for the items were assumed by the experts, which means that they might not represent their real responses. For example, the fourth row in the Q matrix (see Table II) means that the experts assumed that the students should be proficient in *electric charge* and *Coulomb's law* first if they provide the correct answer to item 4 (see Sec. III C). However, the students might give a correct response to item 4 without being proficient in the above concepts, which means the theoretical response processes do not fit the realistic response processes of the students. Therefore, as mentioned above (see Sec. III A), the Q matrix defined by the experts was amended by the reasons provided by the students in the R tier of the instrument. The following examples demonstrate the amendment process.

Table I illustrates that the experts assume that students should be proficient in *electric charge* (misconception 1.6) and *electric field* (misconception 2.4, 2.14, and 2.17) if they give the correct answer to item 3. Many students also used Coulomb's law to calculate the forces between two metal spheres and said "the field is generated by a point electric charge, which Coulomb's law can calculate; then the total field at point D can be calculated using the superposition of an electric field." Therefore, if a student provides a correct answer to this item, they might also be proficient in *Coulomb's law* and the third row of the Q matrix is amended as *(1,1,1,0,0,0)*.

The other items 1, 5, 7, 8, 10, 11, 12, 13, and 16 are also amended through the above process, and the Q matrix illustrated in Table II is the amended Q matrix. This amendment process ensures that the instrument's construct fits the detailed nature of the responses by the students.

D. Confirmation of students' understanding of electrostatics

Students' concept profiles and ability estimates are the diagnostic output results of the instrument [114], including students' overall understanding of all concepts and students' understanding of specific concepts. This subsection details the interpretation of reports that emerged from the instrument and confirms them.

1. Examinee skill profile and ability estimates

Figure 2 aggregates the proficiency status across 1850 students. It demonstrates the estimated skill distributions produced by the G-DINA model, which indicates that each concept has been grasped by less than half of the examinees in the sample (e.g., electric charge and Coulomb's law). This result is consistent with the existing studies that students have above-average difficulties with electrostatics [8].

The G-DINA model estimates 64 possible skill arrangements. Because of the large number of skill profiles, Fig. 3 reports only the estimated skill profiles that account for at least 1% (19 participants) of the sample of students. As illustrated in Fig. 3, most students were classified with nonproficiency in each concept ($\{0,0,0,0,0,0\}$, 42%), and some students were classified as being proficient in several (one, two, or more) concepts. For example, around 21% of the students were classified as mastering electric charge and electric potential (i.e., $\{1,0,0,0,1,0\}$). This result again emphasizes that students have above-average difficulties with electrostatics [8].

Figure 2 illustrates the concept proficiency distribution, and Fig. 3 illustrates the concept profile among the participants. Admittedly, these two results demonstrate concept proficiency across a large cohort of students and the students classified into different profiles, respectively. These two results are formed by the students' concept profiles, demonstrating that a student grasps the concepts. Simultaneously, the student's concept profile can provide more specific diagnostic information relevant to classroom instruction and student learning, expressed in terms of a user-friendly diagnostic report.

The instrument's primary outcome is the student concept profiles, which can demonstrate students' conceptual state in learning *electrostatics*. As mentioned in Sec. II A, six concepts are specified when the instrument is developed to diagnose students' conceptual understanding of electrostatics: electric charge, electric field, Coulomb's law, electrical potential energy, electric potential, and capacitance. A case analysis was conducted to demonstrate the value of students' concept profiles, and four individual diagnostic reports (students A, B, C, and D) are presented in Fig. 4.

In this example, four students received the same overall score (i.e., they gave 12 correct answers to all 20 items) with different concept proficiency decisions and a different proficiency probability for each concept. As mentioned in Templin *et al.* [67], statistically, students with a greater than or equal to a 0.6 probability of proficiency in a concept are classified as proficient in that concept and less than 0.4 as nonproficiency. Students between 0.4 and 0.6 may be classified as "undecided" because the assessment does not provide enough information about them [67,98].

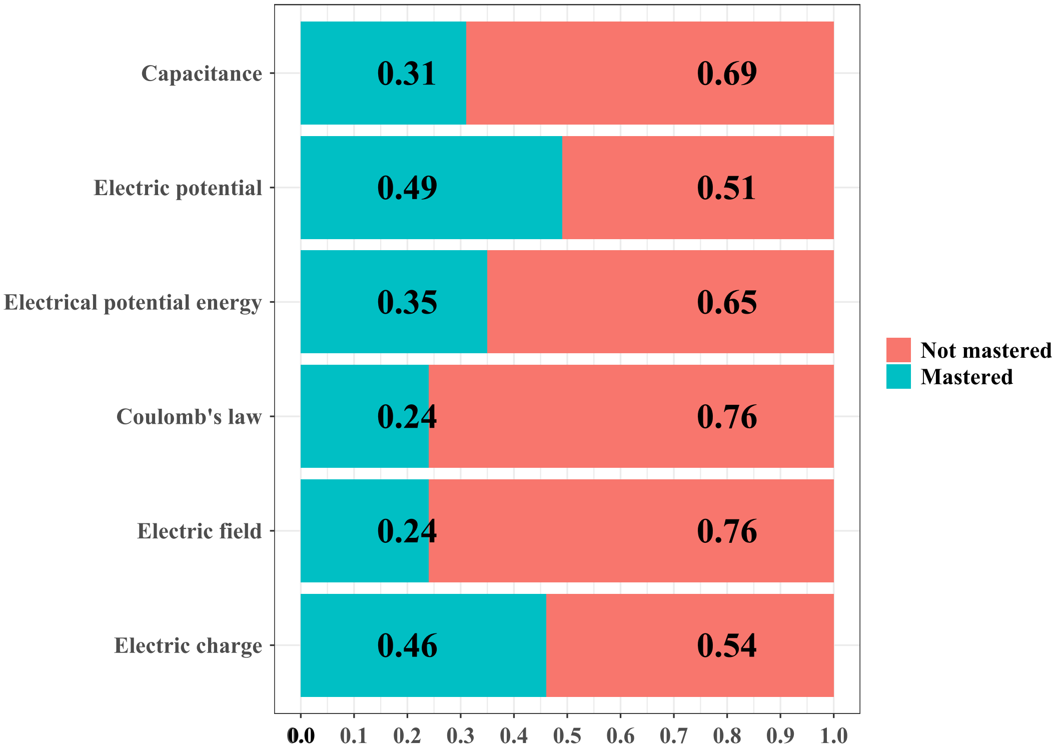

FIG. 2. Concepts proficiency (mastery) distribution.

Figure 4 demonstrates that students with the same observed scores did not necessarily have the same concept profiles; for example, though all students obtained 12 points, student A is proficient in electric charge and electrical potential, but not electric field, Coulomb's law, electric potential energy, and capacitance, while student B is proficient in electric charge, electric field, electrical potential, and capacitance, but not electrical potential energy and Coulomb's law. Students C and D are proficient in different concepts though the total score is equal. If a single observed score was provided to students, it could not inform them about their strengths and weaknesses of conceptual understanding related to electrostatics because it masks fine-grained, specific diagnostic information. For example, there is no difference between the observed score of students A, B, C, and D, but students B, C, and D became proficient in four concepts (with different concepts grasped), while student A became proficient in only two

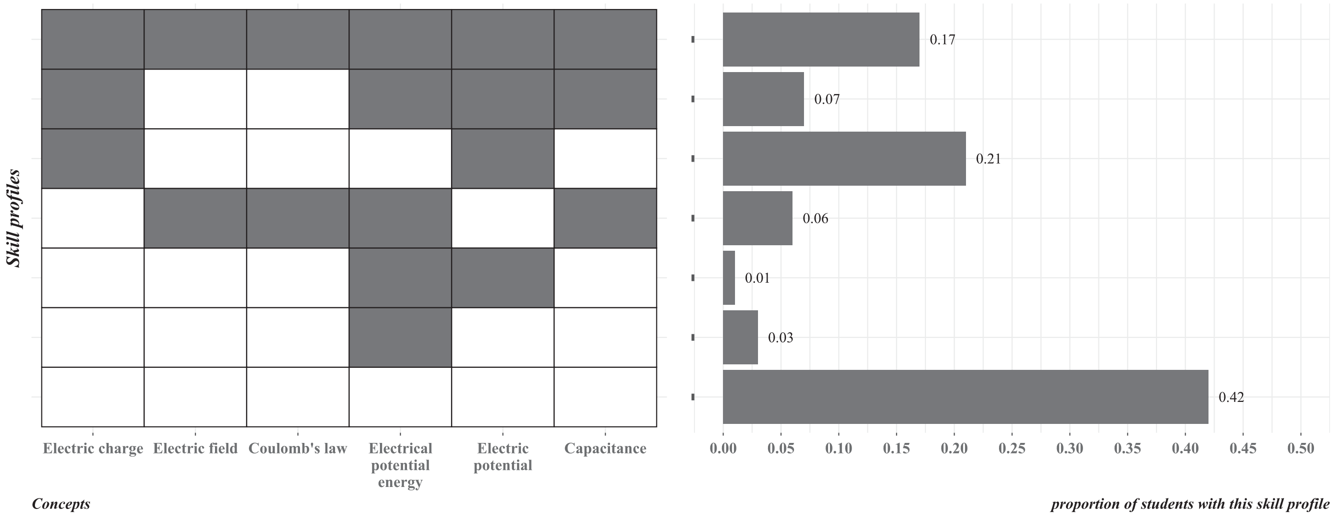

FIG. 3. Concepts profile of participants.

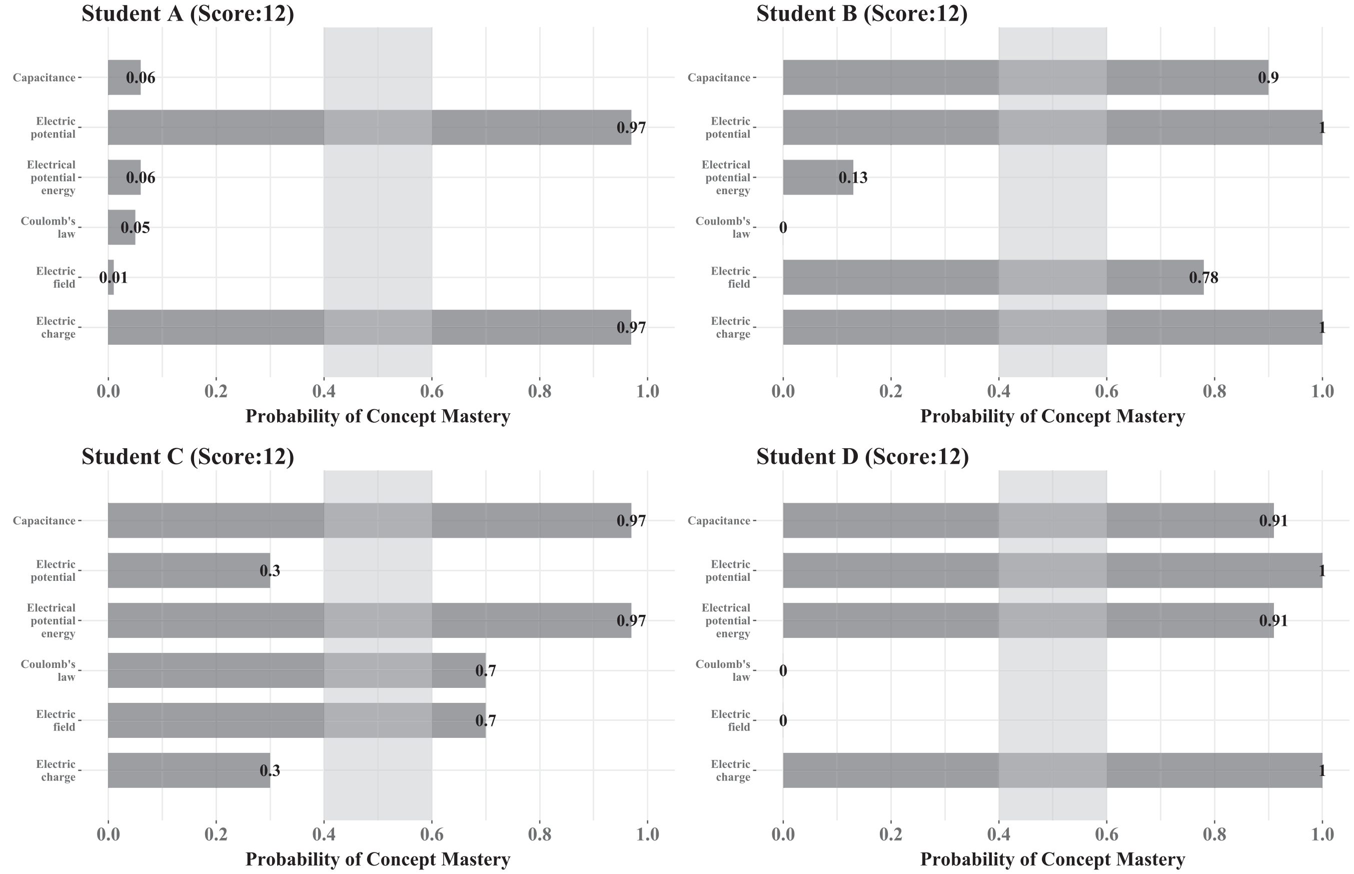

FIG. 4. Example of individual diagnostic reports about students' conceptual understanding related to electrostatics.

concepts. It is also possible that this nonrelationship can be interpreted as highlighting the need for diagnostic concept reports.

2. The reliability of the reports about students' concept profiles

This subsection uses student A as an example to confirm the reliability of the cognitive diagnostic report about the students' conceptual understanding, which might illustrate whether the diagnostic reports illustrated in Fig. 4 represent students' conceptual understanding related to electrostatics.

The reasons for each item provided by student A are summarized, and the results are as follows:

In item 15, student A thought as follows: "the capacitance of the capacitor might change with an insulator inserted between two conductor parallel plates" and concluded that "capacitance will be smaller," which means he might think that "inserting an insulator between two conductor parallel plates reduces the capacitance of the system" (see Table I, misconception 6.3). He also thought that "only the distance traveled of a particle will affect the net work of electric field force" in item 8, which means that he might hold misconception 4.1 from Table I at a minimum. In item 3, he stated, "the charge of these two metal spheres will be divided evenly; then, electric field strength can be calculated through Coulomb's law," which means he does not establish a clear difference between the concepts of field intensity and electric force (see Table I, misconception 2.4). In addition, though this student believes that the electric field strength could be calculated through Coulomb's law, the erroneous computational results might indicate the nonproficiency in Coulomb's law. The above reason provided by this student means he holds some misconceptions about capacitance, electric potential energy, Coulomb's law, and electric field, which supports this student's diagnostic report illustrated in Fig. 4. The reasons provided by the other three students in Fig. 4 also support the diagnostic report provided by the instrument.

In addition, this study also discusses an important source of valid evidence derived from analyses of the relationship of the newly created instrument to external variables [131,132]. As mentioned above, the instrument produces a list of concepts that students might or might not possess, based on the evidence of tasks in the instrument that they perform [69]. As Junker and Sijtsma [133] demonstrated, in general, the more item-relevant concepts a student possesses, the higher the average of the student's ability (labeled as θ, possibly multidimensional) [79]. Therefore, in this study, the relationships between the number of proficient concepts and the student's ability (obtained from the 2PL model, see Table VII) were calculated as Lee and Luna-Bazaldua [79] suggested. The Pearson correlation coefficient r between them is equal to 0.94 (a large effect size when $r > 0.5$, see [134]), which means the instrument ensures a student with a higher ability will become proficient in more concepts.

E. The inference of the secondary concepts' correlations and hierarchies

CDA also can provide the secondary concepts' correlations and hierarchies (see Sec. II D). This can be inferred from the proficiency behavior of students, and a detailed process for this is described as follows: First, the tetrachoric correlations between every pair of concepts related to electrostatics were computed using students' concept profiles. Then an exploratory factor analysis (EFA) was run using data about tetrachoric correlations, and two factors were found. Electric charge, electrical potential energy, and electric potential belong to the first factor, and the electric field, Coulomb's law, and capacitance belong to the other factor. Combined with the percentage of students in the test population possessing the individual concepts, students' construct of electrostatics might be induced. For example, students who become proficient in electric charge often become proficient in electric potential $(0.17 + 0.07 + 0.21 = 0.45)$ in contrast to students who become proficient in electric potential but not electric charge (0.01); therefore, this might indicate that becoming proficient in electric charge could be a prerequisite for becoming proficient in electric potential. Students constructs of electrostatics can be described as follows: The secondary concepts of electrostatics can be attributed to two factors; for the first factor, the electric charge is the prerequisite of electric potential, which is the prerequisite of electric potential energy; and for the second, the electric field is the prerequisite of Coulomb's law (always mastered together) and capacitance. These two factors correlate through the relations between the electric field and electric potential energy. This information can be aggregated in the classroom in schools for instruction planning, low-stakes decision making, or inferring a learning sequence.

VI. DISCUSSION, LIMITATIONS, AND OUTLOOK

The central principle of the instrument developed in this study is that it provides fine-grained results regarding students' cognitive strengths and weaknesses related to the conceptual understating of electrostatics. The development process is organized around two activities, respectively, the development and validation of the instrument.

As for the development step, the secondary concepts (e.g., work and electric field) are defined first; then, the items are designed using their misconceptions. The relationships between the misconceptions and test items are stored in the Q matrix, which is input data for CDA models. In this step, the secondary concepts are defined by educational and didactic experts who have a deeper conceptual understanding than novices, and the misconceptions related to these concepts are summarized from the existing studies. The items are developed by a focus group consisting of a high school physics teacher and a science education researcher rather than only one researcher. The above process guaranteed the improvement of the quality of test content to the maximum extent by allowing the test content to be part of the response data procedure [135].

In the validation step, the best model was selected through three levels of model fit: model, item, and personal level. Additionally, the instrument is validated from item quality, overall test reliability, instrument structure, and diagnostic results. The best model selection ensures the model fits data at the test level, item level, and the level between the test result and students' true conceptual understanding. The later analysis ensures that the diagnostic reports represent students' conceptual understanding reliably and validly.

The above process and the analysis of the validity and reliability evidence of the diagnostic reports confirm that the instrument developed in this study is a discriminant, accurate, and reliable instrument for determining students' conceptual understanding related to electrostatics.

However, there are also limitations in this study. First, the item discrimination from the CDA model in item 7 is lower than 0.30, which means that this item might not achieve the intended classification role of the item. Though researchers in the focus group agreed that students should be proficient in understanding electric fields to answer this item correctly [students believe the change in electric potential also affects the path of the particle; therefore, the seventh row of the Q matrix is amended as *(0,1,0,0,1,0)*, see Table II], they should also be proficient in Newton's law (especially Newton's second law) which might help them to analyze the relationship between force and motion (numerous students answer this item incorrectly due to the lack of knowledge about Newton's law). The instrument aimed to diagnose students' conceptual understanding of electrostatics rather than force and motion; at the same time, this shortcoming cannot prevent or reduce the accurate and reliable proficiency classification. Therefore, item 7 is not removed from the instrument, though it has a lower item discrimination value. Second, the G-DINA model illustrated a higher proportion of spuriously low scores, meaning students become proficient in all acquired attributes but fail to answer the item, indicating underestimated high-level students correctly. Nonetheless, the misfit was considered negligible because the underestimated high-level students were still classified as proficient in all skills. The accuracy and reliability skill proficiency classification values mentioned above confirm this. Third, the test results demonstrate that students have a mean score of 42.4% on the test, and there are many concepts that more than half of the students in the sample have not mastered yet that were detected by the test. This result indicates that the test is rather difficult for students. However, the misconceptions summarized from existing studies are the misconceptions that must be overcome to understand the electrostatics, and the difficulty of the test has not affected the classification consistency and accuracy from the analysis results of reliability (see Table VIII). Therefore, the difficulty of the test can be attributed to students' superficial understanding of concepts related to electrostatics rather than badly designed items. Of course, as Kiviniemi *et al.* [136] suggests that a slightly "too difficult" test may be more optimal to map the gain of students from a pretest to a post-test if the test is still discriminating and reliable enough. Therefore, though this test is quite difficult for students, it is still a useful instrument to measure students' understanding of electrostatics. In conclusion, though there are limitations in this study, our work examined the potential application of the CDA approach to classroom conceptual understanding assessment.

Note also that several items in the instrument are not only described using written text but also are depicted in a visual-graphical way, which are the formats of multiple representations [137] (e.g., question 1, question 2; see the Appendix). This implies that to answer the instrument items correctly, students need to mentally integrate partial information from each representation. In general, the ability to understand, use, transfer, select, and create different domain-specific representations (e.g., text, diagrams representations) is represented competence, which is an important learning prerequisite in physics [138]. Several studies have indicated that students experience significant challenges working with multirepresentation [139]; therefore, students' poor performance on some items (e.g., item 7) might also be partly attributable to low representational competence. Unfortunately, this study was mainly concerned with students' performance on the test due to their conceptual understanding of electrostatics rather than their representational competence. Therefore, to effectively examine the psychometric properties of these items, the impact of students' representational competence on their performance on the test should be analyzed in the future.

Compared to CTT and IRT, which provide a total score, the instrument developed based on CDA can provide the percentage of students in the test population who possess certain combinations of concepts (see Fig. 2), the percentage of students in the test population possessing the individual concepts (see Fig. 3), and the fine-grained size of concept proficiency information (see Fig. 4) which can be integrated as one completed report to issue to students, teachers, and parents to demonstrate students' status of conceptual understanding related to electrostatics. Notably, it might be difficult for practitioners (e.g., teachers) to use the instrument and evaluate it with the CDA method if they are not proficient in statistics. Therefore, an online conceptual understanding assessment tool that concerns the CDA and CDA model needs to be built. When students take the test via the online tool, students, teachers, and parents might receive a completed and timely report. For teachers, the reports may not only help them grasp students' conceptual understanding of electrostatics but also help them make decisions in the classroom.

ACKNOWLEDGMENTS

This work is supported by the Shandong Provincial Natural Science Foundation under Grant No. ZR2020MG058. This paper is the output of a research project funded by the National Natural Science Foundation of China (NSFC) under Grant No. 72004166. The authors declare that they have no conflict of interest. Informed consent was obtained from all individual participants included in the study.

APPENDIX: CONCEPTUAL SURVEY INSTRUMENT TO EVALUATE STUDENTS' UNDERSTANDING OF ELECTROSTATICS

Question 1: *Three neutral metal spheres, electrically insulated from the environment, are touching each other, as illustrated in the figure. A negatively charged glass rod approaches the right spheres, as illustrated, causing them to move apart from each other. The three spheres separate while the rod is close and never touching any sphere. Which of the following is right?*

(A) 3 neutral
(B) 2 neutral and 1 positive
(C) 1 neutral, 1 positive, 1 negative
(D) 2 positively charged and one negatively charged
(E) one positively charged and 2 negatively charged
(F) 2 neutral and 1 negative
(G) Your answer: _______

Q1.A Confidence Rating for Q1
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident

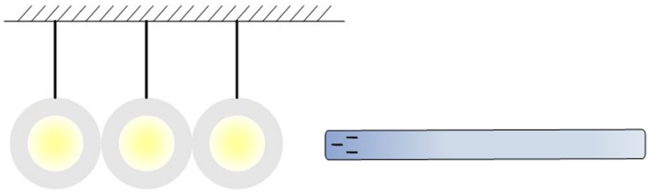

FIG. 5. Figure for Q1.

(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q1.B Scientific reasons for my answer for Q1
(A) Friction is the only cause of static electricity. The glass rod did not touch any spheres; therefore, these three spheres keep electrically neutral. Additionally, there are three neutral spheres.
(B) Your reason: _______

Q1.C Confidence Rating for Q1.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 2: *Two small identical metal spheres on insulated stands carry charges* $(+2q)$ *and* $(+4q)$*, respectively. When the centers of the spheres are separated by a distance* d $(d \gg$ *radius of spheres), one exerts an electrostatic force of magnitude F on the other. The spheres are now made to touch each other and are then brought back to the distance 2d apart. What will be the magnitude of the electrostatic force that one sphere now exerts on the other?*
(A) 2F
(B) 9F/32
(C) F/4
(D) F
(E) Your answer: ________

Q2.A Confidence Rating for Q2
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q2.B Scientific reasons for my answer for Q2
(A) Charges do not transfer between conductors with charges of the same sign; therefore, the charges carried by these two spheres would be maintained. Because the distance between them became 2d, the one exerts an electrostatic force of magnitude F on the other would become half of F.
(B) Charges do not transfer between conductors with charges of the same sign; therefore, the charges carried by these two spheres would be maintained. Because the distance between them became 2d, the one exerts an electrostatic force of magnitude F on the other would become a quarter of F.
(C) Your reason: _______

Q2.C Confidence Rating for Q2.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 3: *Two small identical metal spheres, on insulated stands, carry charges* $(-Q)$ *and* $(+3Q)$*, respectively, and are placed at point A and point B. When the centers of the spheres are separated by a distance* d $(d \gg$ *radius of spheres), one exerts an electrostatic force of magnitude F on the other. The spheres are now made to touch each other and are then brought back to the distance 3d apart. A test charge* $(-q)$ *is placed over point D. What will be the electric field strength at point D?*
(A) 2F/27Q, along with B to C
(B) $(F\sqrt{3})/27Q$, vertically upwards
(C) $(F\sqrt{3})/27$, vertically downwards
(D) F/Q, vertically upwards
(E) F/Q, vertically downwards
(F) $(F\sqrt{3})/27$, vertically upwards
(G) $(F\sqrt{3})/27Q$, vertically downwards

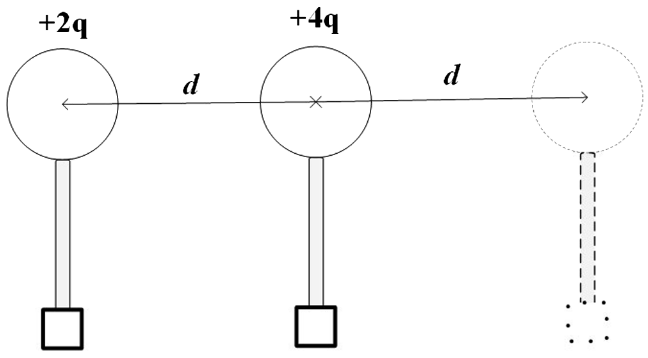

FIG. 6. Figure for Q2.

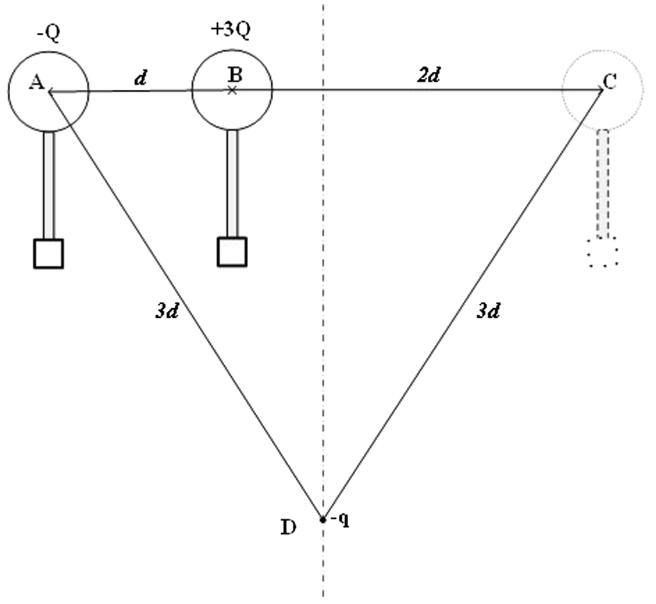

FIG. 7. Figure for Q3.

(H) 2F/27Q, along with C to B
(I) F, vertically upwards
(J) F, vertically downwards
(K) 2F/27, along with C to B
(L) 2F/27, along with B to C
(M) Your answer: ________

Q3.A Confidence Rating for Q3
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q3.B Scientific reasons for my answer for Q3
(A) There would be a transfer between oppositely charged conductors until one of the conductors became neutral, therefore, the left metal sphere might become neutral, and the right metal sphere might be charged by +2Q after losing +Q. The electric field at point D is generated only by the right metal sphere at point C.
(B) Your reason: ______

Q3.C Confidence Rating for Q3.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 4: *Two small metal objects (I and II), each with a net charge of* $+Q$*, exert a force of magnitude F on each other. The right-hand object (II) is now made to touch another metal object whose net charge is* $+5Q$ *and then brought it (II) back to the same distance apart from I. Then, what is the magnitude of the force on the metal objects I and II (the III is withdrawn)?*
(A) F, F
(B) 3F, F
(C) F, 3F
(D) 3F, 3F
(E) Your answer: _______

Q4.A Confidence Rating for Q4
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q4.B Scientific reasons for my answer for Q4
(A) The charges on the two metal objects remain the same after touching; therefore, the I and II have +Q charge, and the magnitude of the force exerted on each other is also equal to F.
(B) Larger “objects” (in charge magnitude) exert larger forces than smaller “objects,” therefore, II exerts 3F on I, and I exert F on II.
(C) Your reason: ________

Q4.C Confidence Rating for Q4.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 5: *As illustrated in the figure, the electric fields were generated by two spheres with* $1\ \mu C$ *charge. Three identical test charges are placed at points A, B, and C. The force of test charges was labeled as* F_A, F_B, *and* F_C. *Which of the following represents the magnitude of the force of three test charges from smallest to largest?*
(A) $F_A = F_B = F_C$
(B) $F_A > F_B > F_C$
(C) $F_C > F_B > F_A$
(D) $F_A > F_C > F_B$
(E) $F_B > F_A > F_C$

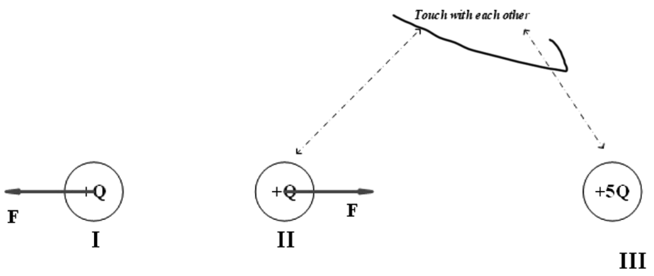

FIG. 8. Figure for Q4.

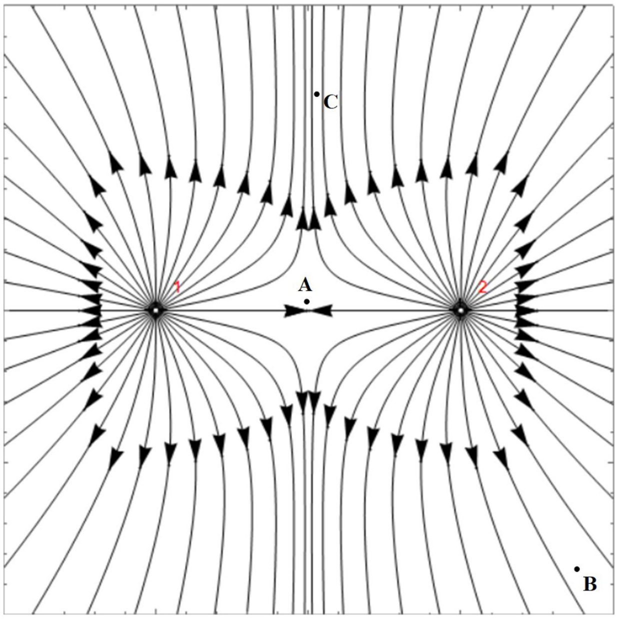

FIG. 9. Figure for Q5.

(F) $F_B > F_C > F_A$
(G) $F_C > F_A > F_B$
(H) Your answer: _______

Q5.A Confidence Rating for Q5
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q5.B Scientific reasons for my answer for Q5
(A) There is no force acting on the test charges because there is no line passing through them; therefore, their forces are all equal to 0.
(B) There was no electric field in gaps between electric field lines; therefore, their forces are all equal to 0.
(C) Your reason: _________

Q5.C Confidence Rating for Q5.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 6: *Four charges are placed on the circumference of a circle of radius 1.0 m and centered at the origin, as illustrated in the figure. What is the magnitude and direction of the electric field strength at the origin (0,0)?*
(A) 0 N/C
(B) 18 000 N/C, and it is at a 45-degree angle with the negative y direction in the third quadrant.
(C) 9000 N/C, and it is at a 45-degree angle with the positive y direction in the first quadrant.
(D) 18 000 N/C, and it is at a 45-degree angle with the positive y direction in the first quadrant.
(E) 9000 N/C, and it is at a 45-degree angle with the negative y direction in the third quadrant.
(F) Your answer: _______

Q6.A Confidence Rating for Q6
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q6.B Scientific reasons for my answer for Q6
(A) There would be no electric field in the absence of an electric charge; therefore, the magnitude of the electric field at the origin is equal to 0.
(B) Your reason: _______

Q6.C Confidence Rating for Q6.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 7: *A particle with +q shot into a region containing two uniform electric fields (Electric field I and electric field II), as illustrated in the figure. The direction of the electric field is as illustrated. Please describe the motion of a particle in the electric field I and select the possible resulting paths for this particle?*

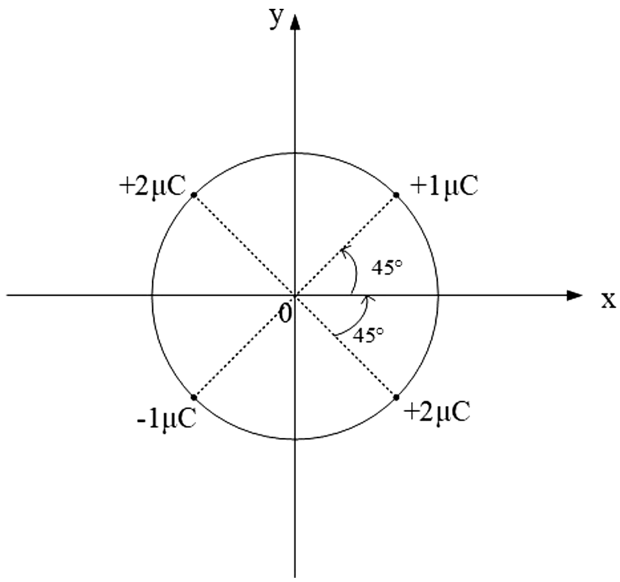

FIG. 10. Figure for Q6.

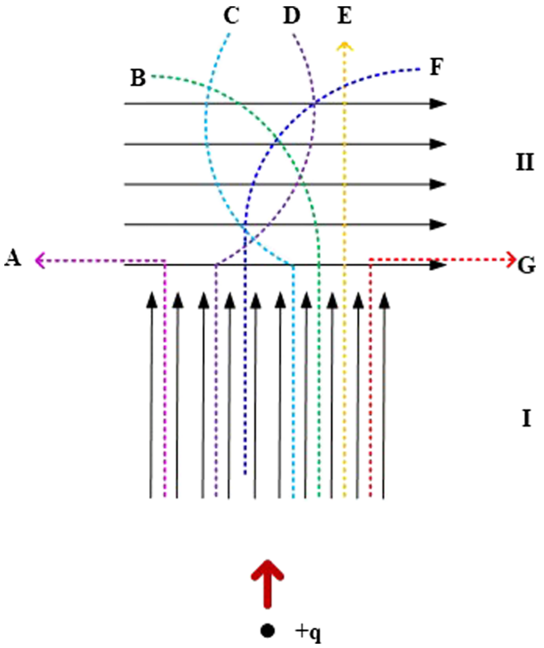

FIG. 11. Figure for Q7.

(A) Uniform motion, A
(B) Uniform motion, B
(C) Uniform motion, C
(D) Uniform motion, D
(E) Uniform motion, E
(F) Uniform motion, F
(G) Uniform motion, G
(H) Uniform variable rectilinear motion, A
(I) Uniform variable rectilinear motion, B
(J) Uniform variable rectilinear motion, C
(K) Uniform variable rectilinear motion, D
(L) Uniform variable rectilinear motion, E
(M) Uniform variable rectilinear motion, F
(N) Uniform variable rectilinear motion, G
(O) Your answer:_______

Q7.A Confidence Rating for Q7

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q7.B Scientific reasons for my answer for Q7

(A) Charged particles in a uniform electric field move with a constant velocity; therefore, this particle might be in uniform motion. Any charged particle independent of the charge polarity moves in the direction of the electric field; therefore, G represents the possible paths for this particle.
(B) Charged particles in a uniform electric field move with a constant velocity; therefore, this particle might be in uniform motion. Electric field lines mean the line of trajectory, and G represents the possible paths for this particle.
(C) Your reason:_________

Q7.C Confidence Rating for Q7.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 8: *As illustrated in the figure, the electric fields were generated by two spheres with* 1-μ*C charge. A certain charge* (+*q*) *is carried from points A to B, C, D, and E. Which of the following statements in relation to the work done by the electric force is correct? Additionally, is the work done by electric force carrying a* + *q charge from A to C positive or negative?*

(A) $W_{AE} > W_{AD} > W_{AC} > W_{AB}$, negative
(B) $W_{AC} = W_{AB} > W_{AD} > W_{AE}$, negative
(C) $W_{AC} = W_{AB} > W_{AD} = W_{AE}$, negative
(D) $W_{AE} > W_{AD} = W_{AC} > W_{AB}$, negative
(E) $W_{AE} > W_{AD} > W_{AC} > W_{AB}$, positive
(F) $W_{AC} = W_{AB} > W_{AD} > W_{AE}$, positive

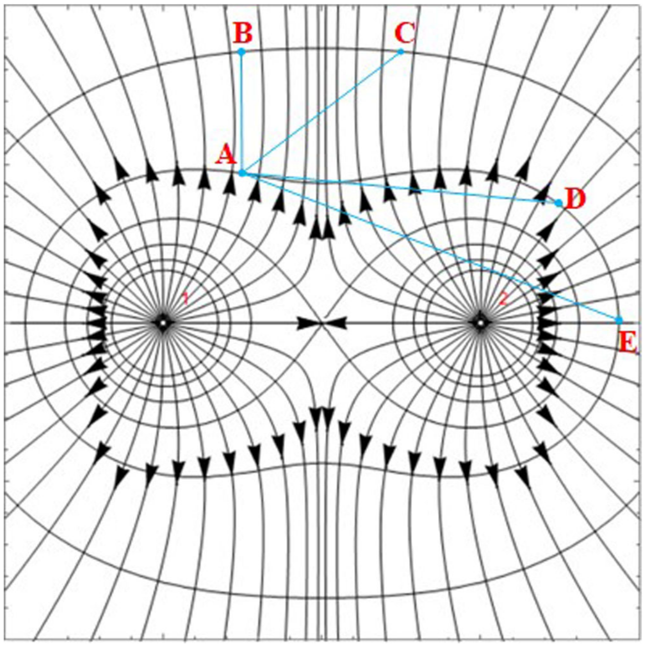

FIG. 12. Figure for Q8.

(G) $W_{AC} = W_{AB} > W_{AD} = W_{AE}$, positive
(H) $W_{AE} > W_{AD} = W_{AC} > W_{AB}$, positive
(I) Your answer: _______

Q8.A Confidence Rating for Q8

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q8.B Scientific reasons for my answer for Q8

(A) The larger the path, the bigger the work is needed; therefore, $W_{AE} > W_{AD} > W_{AC} > W_{AB}$.
(B) The larger the path on the equipotential surface, the bigger the work is needed; therefore, $W_{AE} > W_{AD}$, and $W_{AC} > W_{AB}$.
(C) Your reason: ______

Q8.C Confidence Rating for Q8.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 9: *As illustrated in Fig. 9, the left scale measures the distance of the equipotential line apart from 0 cm. Additionally, the right labels represent the electric potential of each equipotential line. What is the electric field strength at point A?*

(A) Do not know
(B) 500/3 V/m, along with negative y direction
(C) 500/3 V/m, along with positive y direction

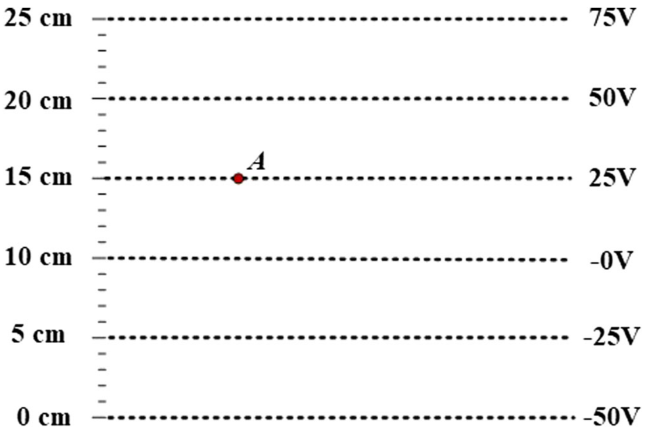

FIG. 13. Figure for Q9.

(D) 500 V/m, along with negative y direction
(E) 500 V/m, along with positive y direction
(F) Your answer: ______

Q9.A Confidence Rating for Q9
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q9.B Scientific reasons for my answer for Q9
(A) The magnitude of electric field strength depends on the force on a charge at point A ($E = F/q$), and there is no charge at point A; therefore, we cannot know the magnitude of electric field strength there.
(B) $E = \frac{\varphi}{x}$, therefore, the magnitude of electric field strength is equal to $\frac{25V}{0.15m} = 500/3$ V/m
(C) Your reason: ______

Q9.C Confidence Rating for Q9.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 10: *As illustrated in Fig. 10, positively charged objects are fixed to the insulation rod (III) through an insulation rod. A neutral metal object hangs in the insulation rod (III) through an insulating rope with an insulation ring. There is no friction between the insulation rod (III) and the insulation ring. Which of the following statements about neutral object (II) and charged object I is correct?*
(A) It(II) will continue staying at the same place.
(B) It(II) might be attracted by object I, then it will touch object I, and the positive charge of object I only have might be evenly divided by object I and object II.
(C) It(II) might be excluded by object I, then it will move away from object I.

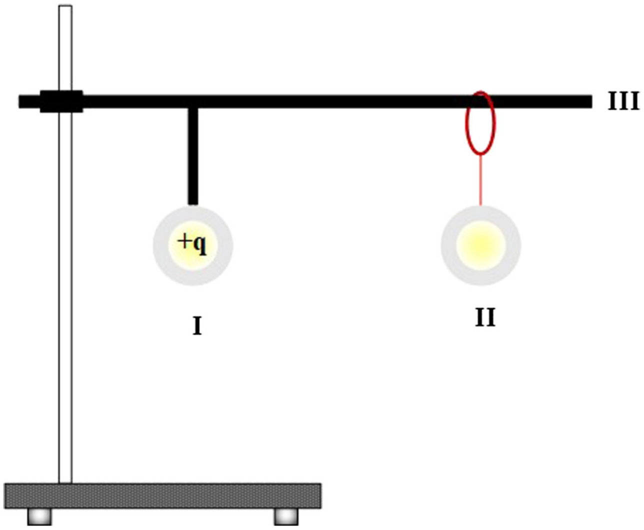

FIG. 14. Figure for Q10.

(D) It(II) might be attracted by object I, then it will touch object I, and the object II will be positively charged.
(E) It(II) might be attracted by object I, then it will touch object I, and the object I will be negatively charged.
(F) It(II) might be attracted by object I, then it will touch object I, and the object II will be negatively charged.
(G) Your answer: _______

Q10.A Confidence Rating for Q10
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q10.B Scientific reasons for my answer for Q10
(A) A neutral object has no charge; therefore, it would not be attracted by object I, and it will continue staying at the same place.
(B) Electrostatic objects cannot attract neutral objects; therefore, they will stay in the same place.
(C) Charged objects have only one type of charge (positive charge), and object I will attract object II; there is no friction between the insulation rod (III) and the insulation ring. Therefore, these two objects will touch each other finally. The positive charge of object I only have might be evenly divided by object I and object II
(D) Your reason: ______

Q10.C Confidence Rating for Q10.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 11: *Experimenter A uses a test charge $+q_0$, and experimenter B uses a test charge $-q_0$ to measure an electric field produced by two parallel plates. A finds a field that is ()*

(A) Greater than the field found by B
(B) The same as the field found by B
(C) Less than the field found by B
(D) Either greater or less than the field found by B, depending on the forces on the test charges
(E) Your answer: _______

Q11.A Confidence Rating for Q11

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q11.B Scientific reasons for my answer for Q11

(A) The magnitude of electric field strength depends on the force on a charge at a point ($E = \frac{F}{q}$), and we do not know the force on $+q_0$ and $-q_0$; therefore, we cannot know the size relations between the electric field that A found and B found.
(B) The electric field would not change when a new positive charge is added to the system, while the field would decrease when a negative charge is added. Therefore, A finds a field that is greater than the field found by B.
(C) Your reason: _______

Q11.C Confidence Rating for Q11.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 12: *As the figure illustrates, the diagram of electric charges induced in conductive objects (shapes) by the electrostatic field (lines with arrows) of a nearby charge (+) due to electrostatic induction. There are some erroneous field lines in this figure. Please observe the pictures carefully and tell me the number of errors in this figure.*

(A) 0
(B) 1
(C) 2
(D) 3
(E) 4
(F) 5
(G) 6
(H) 7
(I) 8
(J) 9
(K) 10
(L) 11
(M) 12
(N) Your answer: ______

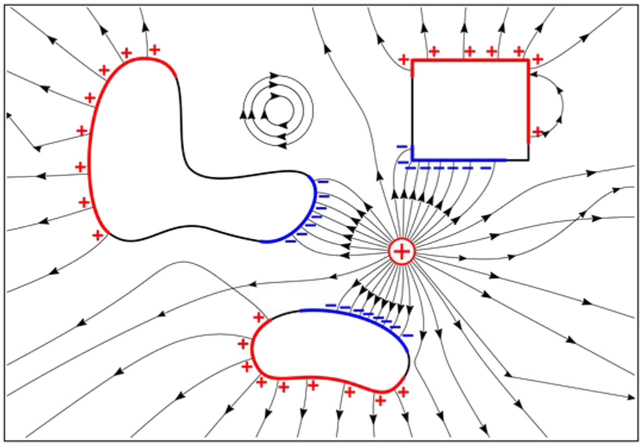

FIG. 15. Figure for Q12.

Q12.A Confidence Rating for Q12

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q12.B Scientific reasons for my answer for Q12

(A) Electric field had volume, force, or density; therefore, electric field lines are real entities, so I can draw electric field lines freely, which means there is no error in this figure.
(B) Your reason: ______

Q12.C Confidence Rating for Q12.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 13: *As the figure illustrates, the electric field of a positive charge, a negative test charge, is placed at point A with the initial velocity v; which of the following statements in relation to the motion of it is correct?*

(A) It will move in a circular motion along the dotted line on which point A is on.
(B) It will move to the positive charge, and its velocity will slow down gradually.
(C) It will move to the positive charge, and its velocity will pick up speed gradually.
(D) It moves along a straight line, pointing to a positive charge at a constant acceleration.
(E) It moves along a straight line opposite to a positive charge at a constant acceleration.
(F) It will be far away from the positive charge, and its velocity will slow down gradually.

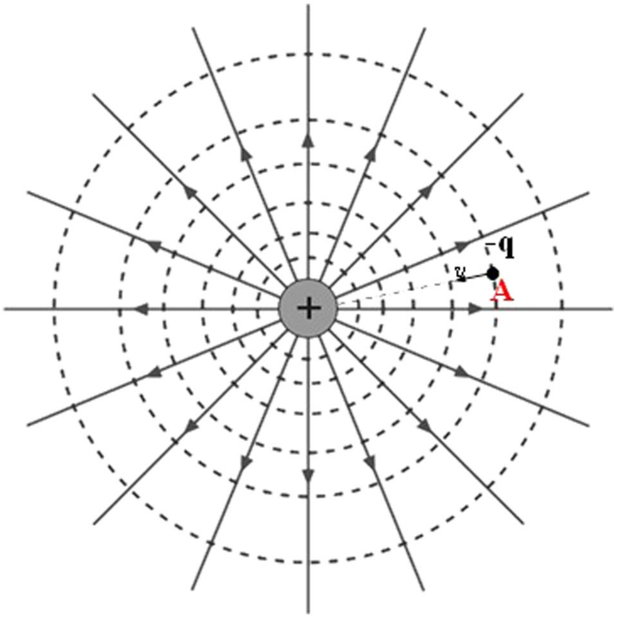

FIG. 16. Figure for Q13.

(G) It will be far away from the positive charge, and its velocity will pick up speed gradually.
(H) It moves along a straight line, pointing to a positive charge at a constant speed.
(I) The movement of it along a continuous curve around the positive charge.
(J) Your answer: ______

Q13.A Confidence Rating for Q13
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q13.B Scientific reasons for my answer for Q13
(A) Particles moving in the opposite direction of the electric field always slow down; therefore, the velocity of it will slow down gradually.
(B) There is no force acting on the charge because no line passes through it; therefore, it moves along the direction of initial velocity, and the magnitude of velocity remains constant.
(C) The dotted lines represent the electric field generated by the positive charge; the direction of its initial velocity is inconsistent with the dotted lines, therefore, the movement of it along a continuous curve around the positive charge.
(D) Your reason: ______

Q13.C Confidence Rating for Q13.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

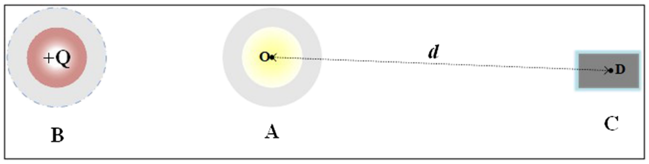

FIG. 17. Figure for Q14.

Question 14: *A neutral metal spheres ball labeled as A, and a metal ball with Q positive charge labeled as B. The A is now made to touch B, and then the B is withdrawn. Then a neutral insulator block is placed at point C, illustrated in the figure. Let us say that the existence of the neutral insulator block will not affect the electric field generated by ball A, and A can be regarded as a point charge. Then which of the following might be the magnitude of the electric field at point D?*
(A) 0
(B) $k \times \frac{2.4\times10^{-19}}{d^2}$
(C) $k \times \frac{2.4\times10^{-19}}{d^3}$
(D) $k \times \frac{2.4\times10^{-19}}{d}$
(E) $k \times \frac{4\times10^{-17}}{d^2}$
(F) $k \times \frac{4\times10^{-17}}{d^3}$
(G) $k \times \frac{4\times10^{-17}}{d}$
(H) Your answer: ______

Q14.A Confidence Rating for Q14
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q14.B Scientific reasons for my answer for Q14
(A) No free charge, so no electrical current means no electrical field inside the insulators. Therefore, the magnitude of the electric field at point D is equal to 0.
(B) Your reason: _______

Q14.C Confidence Rating for Q14.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 15: *The circuit illustrated in the figure consists of a power source, a power switch, and a capacitor. The capacitance of the capacitor is equal to C. The switch is turned on at instant t_1; the switch is turned off at instant t_2. Next, an insulator is inserted between two conductor parallel plates at instant t_3. Finally, the switch is turned on at instant t_4. In this process, capacitance C might vary*

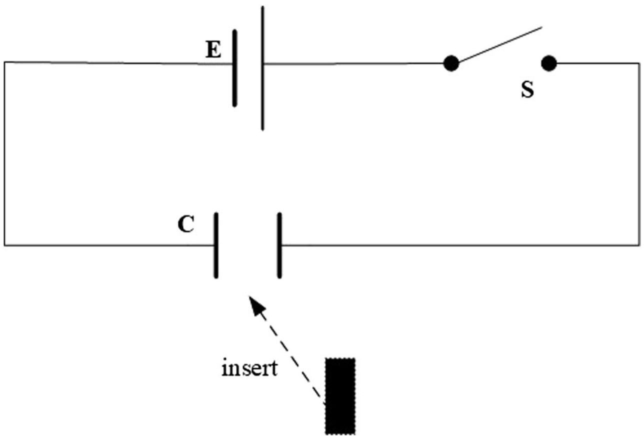

FIG. 18. Figure for Q15.

with different operations. The capacitance C of capacitor at instant t_1, t_2, t_3, t_4 are labeled as C_1, C_2, C_3, C_4. Please decide the size relation between as C_1, C_2, C_3, C_4 and C().

(A) $C_1 = C, C_2 < C, C_3 < C, C_4 < C$
(B) $C_1 = C, C_2 = C, C_3 > C, C_4 > C$
(C) $C_1 = C, C_2 = C, C_3 < C, C_4 < C$
(D) $C_1 = C, C_2 < C, C_3 < C, C_4 > C$
(E) Your answer: ______

Q15.A Confidence Rating for Q15

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q15.B Scientific reasons for my answer for Q15

(A) The capacitance concept has no meaning for uncharged bodies. Therefore, when we turn off the switch, the capacitance gets to 0 when an insulator is inserted between two parallel conductor plates, which means the C will become smaller. Inserting an insulator between two conductor parallel plates reduces the capacitance of the system due to preventing charge transfer from one plate to another, preventing the actual electrical current.
(B) Your reason: ______

Q15.C Confidence Rating for Q15.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 16: *As illustrated in the figure, the electric fields were generated by two spheres with opposite charges, 1-μC charge, and* -1 μC. *Two test charges with $-q$ and $+q$ are placed at points A and B of the two sides of mid-perpendicular symmetrically. Which of the following statements about the size relationships between the electric potential at point A and point B and the electric potential energy of charges at point A and point B is correct?*

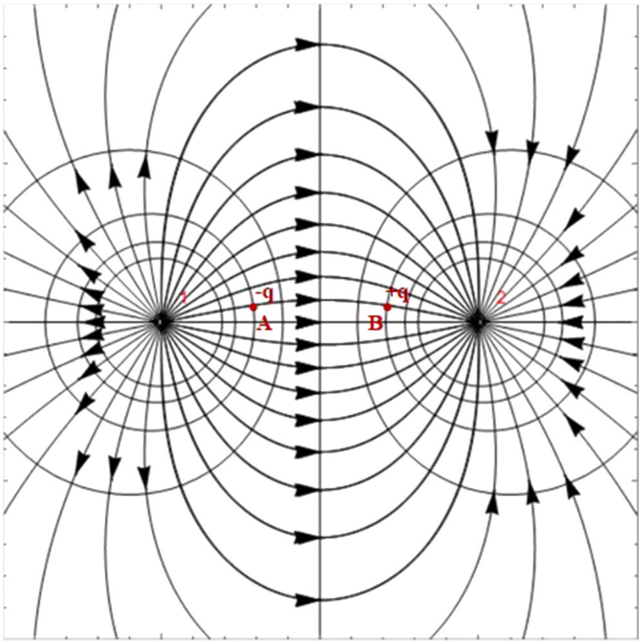

FIG. 19. Figure for Q16.

(A) $\varphi_A < \varphi_B, E_{p-A} < E_{p-B}$
(B) $\varphi_A < \varphi_B, E_{p-A} = E_{p-B}$
(C) $\varphi_A < \varphi_B, E_{p-A} > E_{p-B}$
(D) $\varphi_A > \varphi_B, E_{p-A} < E_{p-B}$
(E) $\varphi_A > \varphi_B, E_{p-A} = E_{p-B}$
(F) $\varphi_A > \varphi_B, E_{p-A} > E_{p-B}$
(G) $\varphi_A = \varphi_B, E_{p-A} < E_{p-B}$
(H) $\varphi_A = \varphi_B, E_{p-A} = E_{p-B}$
(I) $\varphi_A = \varphi_B, E_{p-A} > E_{p-B}$
(J) Your answer: _______

Q16.A Confidence Rating for Q16

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q16.B Scientific reasons for my answer for Q16

(A) Positive charge is associated with high electric potential, and vice versa; therefore, $\varphi_A < \varphi_B$.
(B) Your reason: ______

Q16.C Confidence Rating for Q16.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 17: *As the figure illustrates, the electric field of a negative charge, a positive test charge is placed at point A with the initial velocity v; then it moves from A to B.*

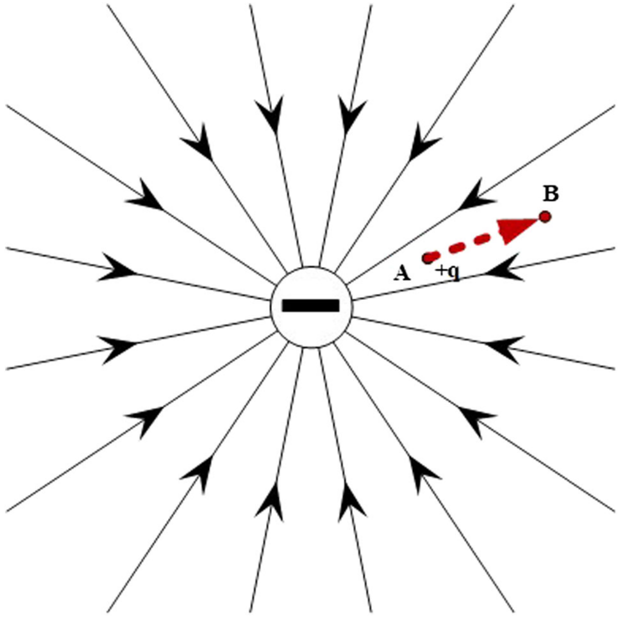

FIG. 20. Figure for Q17.

Which of the following statements in relation to the electric potential energy of it (E_{p-A}, E_{p-B}), and the work done by electric field force W_{AB}, *is correct?*

(A) $E_{p-A} < E_{p-B}, W_A B = 0$
(B) $E_{p-A} < E_{p-B}, W_A B < 0$
(C) $E_{p-A} < E_{p-B}, W_A B > 0$
(D) $E_{p-A} > E_{p-B}, W_A B = 0$
(E) $E_{p-A} > E_{p-B}, W_A B < 0$
(F) $E_{p-A} > E_{p-B}, W_A B > 0$
(G) Your answer: ______

Q17.A Confidence Rating for Q17

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q17.B Scientific reasons for my answer for Q17

(A) Only charged particles on the electric field line are affected by the electric field, and there is no line passing the test charge. Therefore, no force is acting on the test charge, and the work done by the electric field force equals 0.
(B) Your reason: _______

Q17.C Confidence Rating for Q17.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 18: *The figure illustrates three groups of electric potential lines from three electric fields (I, II, and III). Three test charges ($-q$, $+q$, and $-q$) are placed at point A, point B, and point C. Which of the following statements in relation to the magnitude of the electric force exerted on them and the electric potential energy of* E_{p-A}, E_{p-B}, and E_{p-C}, *is correct?*

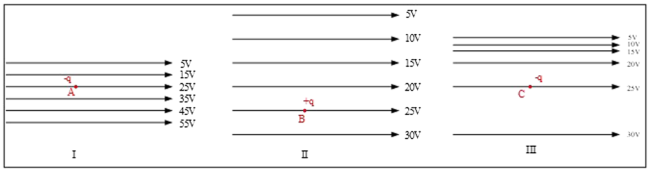

FIG. 21. Figure for Q18.

(A) $F_A < F_B < F_C, E_{p-A} = E_{p-C} < E_{p-B}$
(B) $F_B < F_C < F_A, E_{p-A} = E_{p-C} < E_{p-B}$
(C) $F_B < F_A < F_C, E_{p-A} = E_{p-C} < E_{p-B}$
(D) $F_B < F_C < F_A, E_{p-A} = E_{p-C} < E_{p-B}$
(E) $F_C < F_B < F_A, E_{p-A} = E_{p-C} < E_{p-B}$
(F) $F_C < F_A < F_B, E_{p-A} = E_{p-C} < E_{p-B}$
(G) $F_A < F_B < F_C, E_{p-A} = E_{p-C} > E_{p-B}$
(H) $F_A < F_C < F_C, E_{p-A} = E_{p-C} > E_{p-B}$
(I) $F_B < F_A < F_C, E_{p-A} = E_{p-C} > E_{p-B}$
(J) $F_B < F_C < F_A, E_{p-A} = E_{p-C} > E_{p-B}$
(K) $F_C < F_B < F_A, E_{p-A} = E_{p-C} > E_{p-B}$
(L) $F_C < F_A < F_B, E_{p-A} = E_{p-C} > E_{p-B}$
(M) Your answer: ______

Q18.A Confidence Rating for Q18

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Q18.B Scientific reasons for my answer for Q18

(A) Your reason: ______

Q18.C Confidence Rating for Q18.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 19: *The circuit illustrated in Fig.* 22 *consists of a power source, a power switch, a capacitor, a resistance, an amperemeter, and a voltmeter. If we turn on the switch, which of the following charts might represent the voltage readings and readings on the ammeter from the moment of turning on the switch at time t?*

(A) Your answer: ______

Q19.A Confidence Rating for Q19

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident

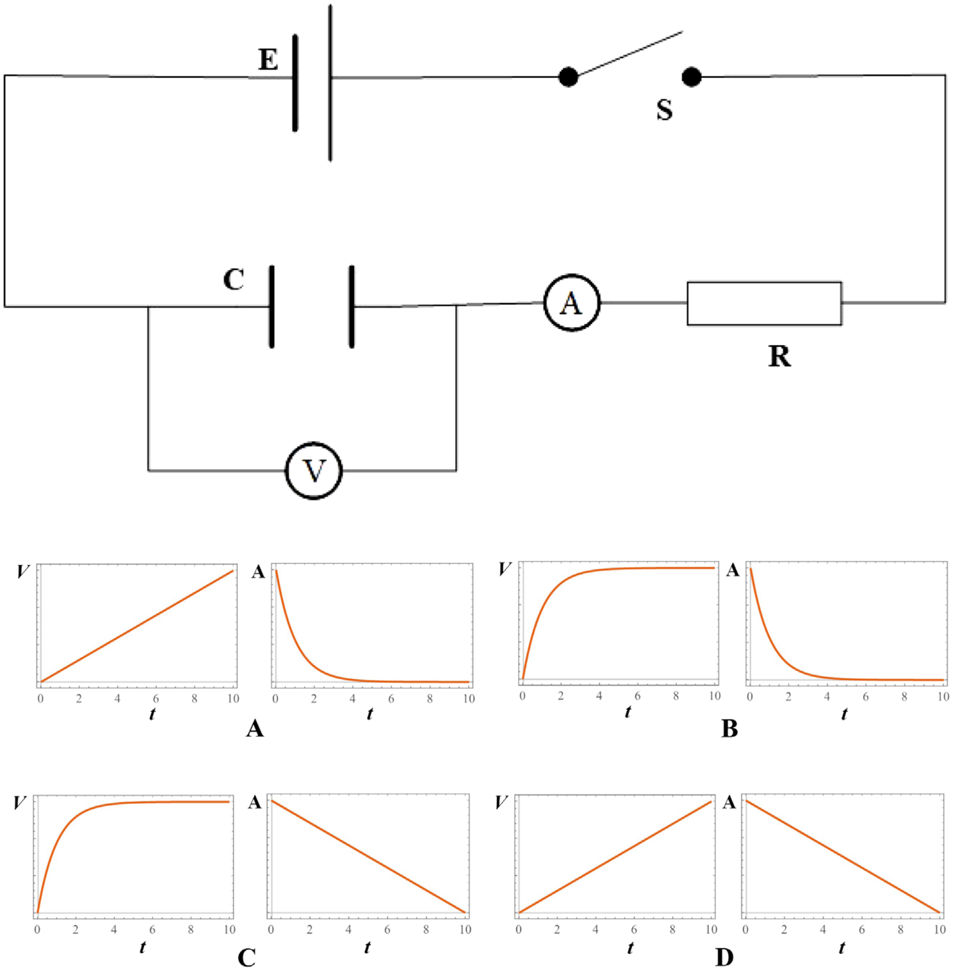

FIG. 22. Figure for Q19.

(E) Very Confident
(F) Absolutely Confident

Q19.B Scientific reasons for my answer for Q19

(A) Your reason: ______

Q19.C Confidence Rating for Q19.B

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

Question 20: *Two small metal objects (I and II), each with a net charge of* $+Q$*, exert a force of magnitude* F *on each other. Which of the following statements about one object exerting force on the other object is correct?*

(A) Mysterious power exerts a force on these two objects.
(B) There is an "action at a distance," and the force is exerted by it.
(C) The electric field exerts forces on these two objects.
(D) A very rarefied and highly elastic substance that permeates all space is the medium of forces radiation, which means this substance is taken as the medium of forces exerted on each object.
(E) Your answer: _____

Q20.A Confidence Rating for Q20

(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

FIG. 23. Figure for Q20.

Q20.B Scientific reasons for my answer for Q20
(A) Your reason: ______
Q20.C Confidence Rating for Q20.B
(A) Just Guessing
(B) Very Unconfident
(C) Unconfident
(D) Confident
(E) Very Confident
(F) Absolutely Confident

---

[1] P. Molinié and S. Boudia, Exhibiting sparks of big science to the public: Electrostatics, atomic machines and experience of paris palais de la découverte, IEEE Trans. Dielectr. Electr. Insul. **16**, 751 (2009).

[2] J. Shen and M. C. Linn, A technology-enhanced unit of modeling static electricity: Integrating scientific explanations and everyday observations, Int. J. Sci. Educ. **33**, 1597 (2011).

[3] A. Benseghir and J.-L. Closset, The electrostatics-electrokinetics transition: Historical and educational difficulties, Int. J. Sci. Educ. **18**, 179 (1996).

[4] J. Park, Analysis of students' processes of confirmation and falsification of their prior ideas about electrostatics, Int. J. Sci. Educ. **23**, 1219 (2001).

[5] C. McMillan III, The role of conceptual reasoning in the solving of textbook problems in electrostatics, Ph.D. thesis, University of Colorado at Boulder, 1990.

[6] A. Barbas and D. Psillos, Evolution of students' reasoning about microscopic processes in electrostatics under the influence of interactive simulations, *Teaching and Learning in the Science Laboratory* (Springer, New York, 2002), pp. 243–254.

[7] W. Chang, Integrating electrostatics with demonstrations and interactive teaching, Am. J. Phys. **79**, 226 (2011).

[8] R. Moynihan, Developing and assessing student's conceptual understanding of electrostatics in upper secondary Physics, Ph.D. thesis, Dublin City University, 2018.

[9] E. Bilal and M. Erol, Investigating students' conceptions of some electricity concepts, Latin-Am. J. Phys. Educ. **3**, 1 (2009), https://dialnet.unirioja.es/servlet/articulo?codigo=3689820.

[10] T. Taşkın and P. Ü. Yavaş, Examining knowledge levels of high school students related to conductors at electrostatic equilibrium and electric field lines using the drawing method, Res. Sci. Educ. **51**, 577 (2021).

[11] E. Adadan, K. E. Irving, and K. C. Trundle, Impacts of multi-representational instruction on high school students' conceptual understandings of the particulate nature of matter, Int. J. Sci. Educ. **31**, 1743 (2009).

[12] S. Kesidou and R. Duit, Students' conceptions of the second law of thermodynamics—an interpretive study, J. Res. Sci. Teach. **30**, 85 (1993).

[13] H.-P. Chang, J.-Y. Chen, C.-J. Guo, C.-C. Chen, C.-Y. Chang, S.-H. Lin, W.-J. Su, K.-D. Lain, S.-Y. Hsu, J.-L. Lin *et al.*, Investigating primary and secondary students' learning of physics concepts in Taiwan, Int. J. Sci. Educ. **29**, 465 (2007).

[14] J. Aguirre and G. Erickson, Students' conceptions about the vector characteristics of three physics concepts, J. Res. Sci. Teach. **21**, 439 (1984).

[15] E. Taslidere, Development and use of a three-tier diagnostic test to assess high school students' misconceptions about the photoelectric effect, Res. Sci. Technol. Educ. **34**, 164 (2016).

[16] C. Guruswamy, M. D. Somers, and R. Hussey, Students' understanding of the transfer of charge between conductors, Phys. Educ. **32**, 91 (1997).

[17] H. D. Brewster, *Electrostatics* (Oxford Book, India, 2009).

[18] N. Jonassen, *Electrostatics* (Springer, New York, 2002).

[19] S. T. Muthiraparampil, Misconceptions in electrostatics among learners at University entry point: A South African case study, Ph.D. thesis, Walter Sisulu University, 2012.

[20] M. Lycoudi, A content analysis of presentations of electrostatics in South African upper secondary school textbooks, Ph.D. thesis, University of the Witwatersrand, 2017.

[21] H. Canbolat, *Electrostatics* (BoD–Books on Demand, Croatia, 2012).

[22] N. Jonassen, *Electrostatics* (Springer Science & Business Media, Denmark, 1998).

[23] B. R. Wilcox, New tools for investigating student learning in upper-division electrostatics, Ph.D. thesis, University of Colorado at Boulder, 2015.

[24] L. Viennot and S. Rainson, Students' reasoning about the superposition of electric fields, Int. J. Sci. Educ. **14**, 475 (1992).

[25] N. Hermita, A. Suhandi, E. Syaodih, A. Samsudin, Isjoni, H. Johan, F. Rosa, R. Setyaningsih, and D. Safitri, Constructing and implementing a four tier test about static electricity to diagnose pre-service elementary school teacher' misconceptions, J. Phys. Conf. Ser. **895**, 012167 (2017).

[26] V. K. Otero, Cognitive processes and the learning of physics part I: The evolution of knowledge from a Vygotskian perspective, *Research on Physics Education* (IOS Press, Amsterdam, 2004), pp. 409–445.

[27] M. A. Siegel and J. A. Lee, "But electricity isn't static": Science discussion, identification of learning issues, and use of resources in a problem-based learning education course, *Proceeding of the Annual Meeting of the National Association for Research in Science Teaching, St. Louis, MO, 2001* (ERIC Clearinghouse, 2001).

[28] A. Yildiz, Discussion on the prospective teachers' understanding level of electric charge, The Turkish Online Journal of Educational Technology 731 (2016).

[29] 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).

[30] M. C. Pocoví, The effects of a history-based instructional material on the students' understanding of field lines, J. Res. Sci. Teach. **44**, 107 (2007).

[31] 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).

[32] N. Tembani, Teaching Newton's second law of motion in Grade 11 Physical Sciences using a Conceptual Change Approach, Ph.D. thesis, University of the Western Cape, 2018.

[33] C. Furió and J. Guisasola, Difficulties in learning the concept of electric field, Sci. Educ. **82**, 511 (1998).

[34] W. Chang, *Utilizing Theme Demonstrations and Concept Maps to Integrate Electrostatics Teaching* (Citeseer, 2007), http://ir.ncue.edu.tw/ir/handle/987654321/16531.

[35] S. Aidoo, A. B. Awuah, and A. O. Mpare, Exploring SHS students' alternative conceptions on electrostatics, Ph.D. thesis, University of Cape Coast, Department of Science and Mathematics Education of the Faculty of Education, 2014.

[36] L. Viennot and S. Rainson, Design and evaluation of a research-based teaching sequence: The superposition of electric field, Int. J. Sci. Educ. **21**, 1 (1999).

[37] S. Rainson, G. Tranströmer, and L. Viennot, Students' understanding of superposition of electric fields, Am. J. Phys. **62**, 1026 (1994).

[38] I. Galili, Mechanics background influences students' conceptions in electromagnetism, Int. J. Sci. Educ. **17**, 371 (1995).

[39] Y. Cao and B.M. Brizuela, High school students' representations and understandings of electric fields, Phys. Rev. Phys. Educ. Res. **12**, 020102 (2016).

[40] B. A. Lindsey, Student reasoning about electrostatic and gravitational potential energy: An exploratory study with interdisciplinary consequences, Phys. Rev. ST Phys. Educ. Res. **10**, 013101 (2014).

[41] L. Doughty, P. van Kampen, and E. McLoughlin, Designing, implementing and assessing guided-inquiry based tutorials in introductory physics, Ph.D. thesis, Dublin City University, Centre for the Advancement of Science and Mathematics Teaching and Learning, 2013.

[42] D. Hammer, Student resources for learning introductory physics, Am. J. Phys. **68**, S52 (2000).

[43] R.L. Hazelton, M.R. Stetzer, P.R. Heron, and P.S. Shaffer, Investigating student ability to apply basic electrostatics concepts to conductors, AIP Conf. Proc. **1513**, 166 (2013).

[44] J. Guisasola, J.L. Zubimendi, J.M. Almudí, and M. Ceberio, Using the processes of electrical charge of bodies as a tool in the assessment of university students' learning in electricity, *Contributions from Science Education Research* (Springer, New York, 2007), pp. 225–236.

[45] J. Guisasola, J.L. Zubimendi, J.M. Almudí, and M. Ceberio, The evolution of the concept of capacitance throughout the development of the electric theory and the understanding of its meaning by university students, Sci. Educ. **11**, 247 (2002).

[46] J. Guisasola, J.L. Zubimendi, and K. Zuza, How much have students learned? research-based teaching on electrical capacitance, Phys. Rev. ST Phys. Educ. Res. **6**, 020102 (2010).

[47] R. Chabay and B. Sherwood, Qualitative understanding and retention, AAPT Announcer **96**, 27 (1997), https://www.physport.org/assessments/assessment.cfm?I=20&A=BEMA.

[48] 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).

[49] R. Rahmawati, N. Rustaman, I. Hamidah, and D. Rusdiana, The development and validation of conceptual knowledge test to evaluate conceptual knowledge of physics prospective teachers on electricity and magnetism topic, J. Pendidikan IPA Indones. **7**, 283 (2018), https://journal.unnes.ac.id/nju/index.php/jpii/article/view/13490.

[50] S. V. Chasteen, R. E. Pepper, M. D. Caballero, S. J. Pollock, and K. K. Perkins, Colorado upper-division electrostatics diagnostic: A conceptual assessment for the junior level, Phys. Rev. ST Phys. Educ. Res. **8**, 020108 (2012).

[51] 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).

[52] Y. Xiao, G. Xu, J. Han, H. Xiao, J. Xiong, and L. Bao, Assessing the longitudinal measurement invariance of the force concept inventory and the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **16**, 020103 (2020).

[53] B. Derek and W. Mark, An introduction to multidimensional measurement using Rasch models, J. Appl. Meas. **4**, 87 (2003), https://nepc.colorado.edu/publication/an-introduction-multidimensional-measurement-using-rasch-models.

[54] J. D. Brown, Classical test theory, *The Routledge Handbook of Language Testing* (Routledge, London, 2013), pp. 337–349.

[55] T. Rusch, P.B. Lowry, P. Mair, and H. Treiblmaier, Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory, Inform. Manage. **54**, 189 (2017).

[56] S. Alagumalai and D.D. Curtis, Classical test theory, *Applied Rasch Measurement: A Book of Exemplars* (Springer, New York, 2005), pp. 1–14.

[57] R.J. de Ayala, *The Theory and Practice of Item Response Theory*, 2nd ed., Methodology in the Social Sciences (The Guilford Press, New York, 2022).

[58] R. De Ayala, Item response theory and Rasch modeling, *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (Routledge, London, 2018), pp. 145–163.

[59] M. H. Trevor, G. Bond, and Zi Yan, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 4th ed. (Routledge, London, 2020).

[60] L. Ding, Seeking missing pieces in science concept assessments: Reevaluating the Brief Electricity and Magnetism Assessment through Rasch Analysis, Phys. Rev. ST Phys. Educ. Res. **10**, 010105 (2014).

[61] M. von Davier, Rasch model, in *Handbook of Item Response Theory, Volume One*, edited by W. J. van der Linden (Chapman and Hall/CRC, Boca Raton, 2016), pp. 31–48.

[62] 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).

[63] J. Hansen and J. Stewart, Multidimensional item response theory and the brief electricity and magnetism assessment, Phys. Rev. Phys. Educ. Res. **17**, 020139 (2021).

[64] M. T. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).

[65] J. L. Snyder, An investigation of the knowledge structures of experts, intermediates and novices in physics, Int. J. Sci. Educ. **22**, 979 (2000).

[66] J. E. Sieber, Problem solving behavior of teachers as a function of conceptual structure, J. Res. Sci. Teach. **2**, 64 (1964).

[67] J. Templin, R. A. Henson *et al.*, *Diagnostic Measurement: Theory, Methods, and Applications* (Guilford Press, New York, 2010).

[68] L. C. McDermott and P. S. Shaffer, Research as a guide for curriculum development: An example from introductory electricity. part I: Investigation of student understanding, Am. J. Phys. **60**, 994 (1992).

[69] Y.-S. Lee, J. de la Torre, and Y. S. Park, Relationships between cognitive diagnosis, CTT, and IRT indices: An empirical investigation, Asia Pac. Educ. Rev. **13**, 333 (2012).

[70] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).

[71] S. Kalyuga, Rapid cognitive assessment of learners' knowledge structures, Learn. Instr. **16**, 1 (2006).

[72] A. C. George and A. Robitzsch, Cognitive diagnosis models in R: A didactic, Quant. Methods Psychol. **11**, 189 (2015).

[73] J. De La Torre, Dina model and parameter estimation: A didactic, J. Educ. Behav. Stat. **34**, 115 (2009).

[74] K. K. Tatsuoka, Analysis of errors in fraction addition and subtraction problems. Final Report, Technical Report No. ED257665, University of Illinois, Urbana-Champaign, 1984, https://eric.ed.gov/?id=ED257665.

[75] M. Kabiri, M. Ghazi-Tabatabaei, A. Bazargan, M. Shokoohi-Yekta, and K. Kharrazi, Diagnosing competency mastery in science: An application of GDM to TIMSS 2011 data, Appl. Meas. Educ. **30**, 27 (2017).

[76] J. De La Torre, L. A. van der Ark, and G. Rossi, Analysis of clinical data from a cognitive diagnosis modeling framework, Meas. Eval. Couns. Dev **51**, 281 (2018).

[77] J. Stewart, F. N. Finley, and W. L. Yarroch, Science content as an important consideration in science education research, J. Res. Sci. Teach. **19**, 425 (1982).

[78] D. Ifenthaler, I. Masduki, and N. M. Seel, The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time, Instr. Sci. **39**, 41 (2011).

[79] Y.-S. Lee and D. A. Luna-Bazaldua, How to conduct a study with diagnostic models, *Handbook of Diagnostic Classification Models* (Springer, New York, 2019), pp. 525–545.

[80] S. Y. Lee, Growth curve cognitive diagnosis models for longitudinal assessment (Order No. 10621005), Ph.D. thesis, University of California, Berkeley, 2018.

[81] A. Mirzaei, M. H. Vincheh, and M. Hashemian, Retrofitting the IELTS reading section with a general cognitive diagnostic model in an iranian EAP context, Stud. Educ. Evaluation **64**, 100817 (2020).

[82] A. C. George, A. Robitzsch, T. Kiefer, J. Groß, and A. Ünlü, The R package CDM for cognitive diagnosis models, J. Stat. Softw. **74**, 1 (2016).

[83] A. E. Lawson and L. D. Thompson, Formal reasoning ability and misconceptions concerning genetics and natural selection, J. Res. Sci. Teach. **25**, 733 (1988).

[84] K. K. Tatsuoka, Rule space: An approach for dealing with misconceptions based on item response theory, J. Educ. Measure. **20**, 345 (1983).

[85] G. Xu and S. Zhang, Identifiability of diagnostic classification models, Psychometrika **81**, 625 (2016).

[86] M. Kim, Y. Cheong, and J. Song, The meanings of physics equations and physics education, J. Korean Phys. Soc. **73**, 145 (2018).

[87] M. Galesic and M. Bosnjak, Effects of questionnaire length on participation and indicators of response quality in a web survey, Publ. Opin. Q. **73**, 349 (2009).

[88] I. S. Caleon and R. Subramaniam, Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions, Res. Sci. Educ. **40**, 313 (2010).

[89] K. G. Tobin and W. Capie, The development and validation of a group test of logical thinking, Educ. Psychol. Meas. **41**, 413 (1981).

[90] D. F. Treagust, Development and use of diagnostic tests to evaluate students' misconceptions in science, Int. J. Sci. Educ. **10**, 159 (1988).

[91] D. Kaltakci-Gurel, A. Eryilmaz, and L. C. McDermott, Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics, Res. Sci. Technol. Educ. **35**, 238 (2017).

[92] H. O. Arslan, C. Cigdemoglu, and C. Moseley, A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain, Int. J. Sci. Educ. **34**, 1667 (2012).

[93] S. J. McKelvie, Graphic rating scales—How many categories?, Br. J. Psychol. **69**, 185 (1978).

[94] https://www.wjx.cn/vj/eaBFKyi.aspx.

[95] Y. Cui, M. J. Gierl, and H.-H. Chang, Estimating classification consistency and accuracy for cognitive diagnostic assessment, J. Educ. Measure. **49**, 19 (2012).

[96] J. De La Torre and J. A. Douglas, Higher-order latent trait models for cognitive diagnosis, Psychometrika **69**, 333 (2004).

[97] J. De La Torre, The generalized dina model framework, Psychometrika **76**, 179 (2011).

[98] S. M. Hartz, A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality., Ph.D. thesis, ProQuest Information & Learning, 2002.

[99] J. L. Templin and R. A. Henson, Measurement of psychological disorders using cognitive diagnosis models, Psychol. Methods **11**, 287 (2006).

[100] A. A. Rupp and J. L. Templin, Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art, Measurement **6**, 219 (2008).

[101] C.-L. Hsu and W.-C. Wang, Variable-length computerized adaptive testing using the higher order DINA model, J. Educ. Measure. **52**, 125 (2015).

[102] S. Hartz and L. Roussos, The fusion model for skills diagnosis: Blending theory with practicality, ETS Res. Rep. Ser. **2008**, i (2008).

[103] S. A. Culpepper and Y. Chen, Development and application of an exploratory reduced reparameterized unified model, J. Educ. Behav. Stat. **44**, 3 (2019).

[104] W. Stout, R. Henson, L. DiBello, and B. Shear, The reparameterized unified model system: A diagnostic assessment modeling approach, *Handbook of Diagnostic Classification Models* (Springer, New York, 2019), pp. 47–79.

[105] W. Ma, C. Iaconangelo, and J. de la Torre, Model similarity, model selection, and attribute classification, Appl. Psychol. Meas. **40**, 200 (2016).

[106] X. Wu, R. Wu, H.-H. Chang, Q. Kong, and Y. Zhang, International comparative study on PISA mathematics achievement test based on cognitive diagnostic models, Front. Psychol. **11**, 2230 (2020).

[107] P.-W. Lei and H. Li, Performance of fit indices in choosing correct cognitive diagnostic models and q-matrices, Appl. Psychol. Meas. **40**, 405 (2016).

[108] N. Jorion, B. D. Gane, K. James, L. Schroeder, L. V. DiBello, and J. W. Pellegrino, An analytic framework for evaluating the validity of concept inventory claims, J. Eng. Educ. **104**, 454 (2015).

[109] M. Arican and O. Kuzu, Diagnosing preservice teachers' understanding of statistics and probability: Developing a test for cognitive assessment, Int. J. Sci. Math. Educ. **18**, 771 (2020).

[110] R. Henson, L. DiBello, and B. Stout, A generalized approach to defining item discrimination for DCMS, Measurement **16**, 18 (2018).

[111] S. Sinharay and M. S. Johnson, Measures of agreement: Reliability, classification accuracy, and classification consistency, *Handbook of Diagnostic Classification Models* (Springer, New York, 2019), pp. 359–377.

[112] M. A. Sorrel, J. Olea, F. J. Abad, J. de la Torre, D. Aguado, and F. Lievens, Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models, Organ. Res. Meth. **19**, 506 (2016).

[113] X. Wu, R. Wu, Y. Zhang, D. Arthur, and H.-H. Chang, Research on construction method of learning paths and learning progressions based on cognitive diagnosis assessment, Assess. Educ. **28**, 657 (2021).

[114] R. Liu, A. C. Huggins-Manley, and O. Bulut, Retrofitting diagnostic classification models to responses from IRT-based assessment forms, Educ. Psychol. Meas. **78**, 357 (2018).

[115] A. Robitzsch, T. Kiefer, A. C. George, and A. Ünlü, CDM: Cognitive Diagnosis Modeling, R package version 7.5-15, 2022, https://CRAN.R-project.org/package=CDM.

[116] J. De la Torre and Y.-S. Lee, Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis, J. Educ. Measure. **50**, 355 (2013).

[117] J. Chen, J. de la Torre, and Z. Zhang, Relative and absolute fit evaluation in cognitive diagnosis modeling, J. Educ. Measure. **50**, 123 (2013).

[118] Wenchao Ma and Jimmy de la Torre, GDINA: The generalized DINA model framework, r package version 2.7, 2019.

[119] O. Kunina-Habenicht, A. A. Rupp, and O. Wilhelm, A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models, Studies in Educational Evaluation **35**, 64 (2009).

[120] O. Kunina-Habenicht, A. A. Rupp, and O. Wilhelm, The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models, J. Educ. Measure. **49**, 59 (2012).

[121] Y. Liu, J. A. Douglas, and R. A. Henson, Testing person fit in cognitive diagnosis, Appl. Psychol. Meas. **33**, 579 (2009).

[122] A. Maydeu-Olivares, Goodness-of-fit assessment of item response theory models, Measurement **11**, 71 (2013).

[123] W. Ma, Evaluating the fit of sequential G-DINA model using limited-information measures, Appl. Psychol. Meas. **44**, 167 (2019).

[124] D. A. F. Robert and L. Ebel, *Essentials of Educational Measurement* (Prentice-Hall, Englewood Cliffs, NJ, 1991).

[125] A. Oosterhof, *Classroom Applications of Educational Measurement* (Macmillan, New York, NY, 2001).

[126] H. Affandy, D. A. Nugraha, S. N. Pratiwi, and C. Cari, Calibration for instrument argumentation skills on the subject of fluid statics using item response theory, J. Phys. Conf. Ser. **1842**, 012032 (2021).

[127] F. B. Baker, *The Basics of Item Response Theory* (ERIC, USA, 2001).

[128] M. S. Johnson and S. Sinharay, Measures of agreement to assess attribute-level classification accuracy and consistency for cognitive diagnostic assessments, J. Educ. Measure. **55**, 635 (2018).

[129] R. J. Adams, Reliability as a measurement design effect, Stud. Educ. Evaluation **31**, 162 (2005).

[130] M. Kane and R. Mislevy, Validating score interpretations based on response processes, *Validation of Score Meaning for the Next Generation of Assessments: The Use of Response Processes* (Routledge, New York, 2017), pp. 11–24.

[131] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing* (American Educational Research Association, Washington, 2014).

[132] D. B. McCoach, R. K. Gable, and J. P. Madura, Evidence based on relations to other variables: Bolstering the empirical validity arguments for constructs, *Instrument Development in the Affective Domain* (Springer, New York, 2013), pp. 209–248.

[133] B. W. Junker and K. Sijtsma, Cognitive assessment models with few assumptions, and connections with nonparametric item response theory, Appl. Psychol. Meas. **25**, 258 (2001).

[134] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge, London, 1988).

[135] F. Davidson, Why is cognitive diagnosis necessary? A reaction, Lang. Assess. Q. **7**, 104 (2010).

[136] T. Kiviniemi, P.-O. Eggen, J. Persson, B. Hafskjold, and E. Jacobsen, Development of a chemistry concept inventory for general chemistry students at Norwegian and Finnish universities, in *Proceedings of Nordic Research Symposium on Science Education* (Norwegian University of Science and Technology, Norway, 2018).

[137] M. Opfermann, A. Schmeck, and H. E. Fischer, Multiple representations in physics and science education—why should we use them?, *Multiple Representations in Physics Education* (Springer, New York, 2017), pp. 1–22.

[138] S. Küchemann, S. Malone, P. Edelsbrunner, A. Lichtenberger, E. Stern, R. Schumacher, R. Brünken, A. Vaterlaus, and J. Kuhn, Inventory for the assessment of representational competence of vector fields, Phys. Rev. Phys. Educ. Res. **17**, 020126 (2021).

[139] M. Stieff, M. Hegarty, and G. Deslongchamps, Identifying representational competence with multi-representational displays, Cognit. Instr. **29**, 123 (2011).