

Development and validation of a two-tier test instrument on simple electric circuits

Benjamin Groß^{1,*}, Lana Ivanjek², Salome Flegel³, Judith Glaesser⁴,
Augustin Kelava⁴, and Jan-Philipp Burde^{1,†}

¹*University of Tübingen, Physics Education Research Group,
Auf der Morgenstelle 14, 72076, Tübingen, Germany*

²*JKU Linz, School of Education, Altenbergerstraße 68, 4040 Linz, Austria*

³*TUD Dresden University of Technology, Chair of Physics Education,
Helmholtzstraße 10, 01069, Dresden, Germany*

⁴*University of Tübingen, Methods Center, Haufferstraße 11, 72076, Tübingen, Germany*



(Received 25 July 2025; accepted 2 December 2025; published 21 January 2026)

In this article, we report the development and validation of a new two-tier concept inventory, the Basic Electricity Assessment of Students' Thinking (BEAST), designed to assess common conceptual difficulties (“student ideas,” such as current consumption and sequential reasoning) and the understanding of fundamental electric circuit concepts (e.g., current, voltage, and resistance) among undergraduate students. The final version of the BEAST covers six student ideas, each linked to a distinct physical concept. Each student idea is assessed using three two-tier single-choice items. To refine the items and evaluate their content validity, including the correctness of answers and the intended categorization of student ideas based on response patterns, feedback and ratings from five experts were incorporated. The test was administered to a sample of $N = 176$ students enrolled in introductory physics courses across six different German universities. Quantitative evaluation using classical test theory and structural analyses demonstrated strong evidence supporting the test's content and construct validity. Confirmatory factor analysis utilizing parsimonious bifactor models showed a good fit of the model to the data.

DOI: [10.1103/mf6b-5xcs](https://doi.org/10.1103/mf6b-5xcs)

I. INTRODUCTION

Electric phenomena and appliances powered by electrical energy are ubiquitous in everyday life. Developing innovative approaches for generating electricity from renewable sources and enhancing its efficient use are central challenges of the 21st century. The societal relevance of this topic is reflected in numerous national high school curricula across Europe—for example, in Germany [1] (p. 16), Austria [2] (p. 111), the United Kingdom at both the primary [3] (p. 23), and secondary levels [4] (p. 12), France [5] (p. 102), Sweden [6] (pp. 174–177), and also by the widely adopted NGSS in the United States [7]. In Germany, for instance, seventh- or eighth-grade students spend approximately ten weeks studying electricity [8]. However, despite substantial teaching efforts and the development of new curricula using various approaches in the past decades [9–11], many students struggle to develop an adequate understanding of even simple electric

circuits and instead hold a number of ideas that are problematic from a scientific point of view. This lack of conceptual understanding has been shown to persist even among undergraduate university students and in-service teachers of physics [12–16]. Over recent decades, various test instruments [17–20] have been developed to assess students' conceptual understanding of circuits. In this article, we analyze the properties of those previously developed test instruments and derive development goals for our new concept inventory. Subsequently, we describe its development process and evaluate its psychometric properties, i.e., reliability scores and validity analyses.

II. BACKGROUND

A. Subject: Simple electric circuits

In this paper, we define simple electric circuits as circuits driven by a dc source—i.e., a source of constant voltage, like a battery—and containing only basic elements like resistors, light bulbs, and switches connected by wires as well as instruments for measuring current or voltage. We characterize the simplicity of these circuits by restricting their complexity primarily to series or parallel circuits containing no more than three of the aforementioned basic elements. Moreover, we assume all circuit components to be ideal; wires, batteries, and measuring instruments are thus considered to have zero (internal) resistance. This

*Contact author: benjamin.gross@uni-tuebingen.de

†Contact author: jan-philipp.burde@uni-tuebingen.de

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

TABLE I. Overview of student ideas assessed by the BEAST test instrument.

Concept	Student idea	Abbreviation
Current	Current consumption	CC
Battery	Batteries as a source of constant current	BSCC
Voltage	Distinguishing between current and voltage (“No voltage without a current”)	VI
Electric circuit as a system	Sequential reasoning	SEQ
Resistance	Parallel and series circuits—Focusing on number of resistors instead of configuration	RES
Energy	Focusing on current alone when determining energy dissipated by a resistor	EN

idealization allows for a purely conceptual analysis of the circuits without necessitating numerical calculations.

B. Student thinking about scientific concepts

Since the 1970s, there has been an extensive amount of research on students’ (and teachers’) conceptions regarding scientific topics, with significant attention directed toward electricity and electric circuits specifically [21]. Various terminologies such as “misconceptions,” “alternative frameworks” [22], “phenomenological primitives” [23], and “student ideas” [24] have been introduced to describe the conceptual difficulties that students encounter in scientific reasoning. Each of these terms carries slightly distinct connotations, reflecting ongoing debates about the most accurate way to model student thinking. For example, the “framework theory” [25] describes student thinking as governed by a theorylike set of deeply rooted, internally coherent and stable cognitive frameworks established from early childhood. Within this model, the term misconception applies specifically to recurring flawed reasoning patterns observed consistently across various contexts. In contrast, the “knowledge in pieces” perspective views student thinking as guided by generic rules and experience from everyday life, so-called phenomenological primitives or *p primis* [26], and therefore as highly context dependent. Within this perspective, labeling student reasoning as misconceptions would be inaccurate. Previous research on electric circuits indicates substantial inconsistency in students’ responses and reasoning [18,20] making the classification of such reasoning as misconceptions questionable. Therefore, in this paper, we adopt the neutral term student ideas, following Hull *et al.* [27].

Regardless of the theoretical perspective adopted, there is a widespread consensus among education researchers today that learning processes should be understood from a constructivist point of view. This perspective emphasizes that students’ understanding of scientific concepts evolves based on their existing ideas and experiences in the natural world and the classroom. Thus, recognizing and addressing student ideas is considered essential in physics teaching [28]. In summary, student ideas can act as both obstacles and starting points for learning, highlighting the importance and relevance of their reliable assessment.

C. Student ideas about electric circuits

As highlighted above (see also Ref. [21]), student thinking on electricity and electric circuits has been extensively investigated, resulting in numerous documented student ideas. Given the breadth of existing literature on the topic, a comprehensive summary exceeds the scope of this article. Therefore, in this section, we provide a targeted overview of those student ideas on simple electric circuits that are assessed by our test instrument. Table I shows these student ideas, a brief description and the basic concepts of electricity and electric circuits with which they are most closely associated. In the subsequent sections, we refer to the respective categories and test items using the abbreviations given in the table.

1. Current consumption

Students often think of electric current as a substance that is stored in batteries, analogous to oil in a barrel, which is “used up” when powering connected devices such as lightbulbs. Although this idea of consumption manifests in various contexts, we anticipate the so-called attenuation model to be particularly prevalent within our target group. This is because research has shown that following formal instruction at school, most students recognize the necessity of a closed circuit [29]. The attenuation model, however, acknowledges a closed-loop current but incorrectly suggests that circuit components progressively weaken or consume the current, resulting in reduced current after components like resistors or bulbs compared to before them [30,31].

2. Battery as a source of constant current

Research indicates that older high school students, undergraduate university students, and even in-service physics teachers frequently hold the idea that an ideal battery always supplies a constant current [14,15,32] rather than maintaining a constant potential difference across its terminals. This student idea is most effectively revealed in situations where an initial, given circuit is changed, e.g., by adding an additional resistor in a series or parallel circuit without changing the battery. Students frequently predict incorrectly that such modifications would not affect the current in the main branch [14,33].

3. Distinguishing between voltage and current

Students often struggle to recognize electric voltage as an independent concept [17], which may be due to the widespread curricular emphasis on electric current rather than on voltage [14]. They also do not recognize the cause-and-effect relationship between voltage and current but rather see voltage as a property of current [34]. This manifests in the student idea that there can be no voltage in a circuit while there is no current. For example, students commonly assert that no voltage exists across an open switch in a simple circuit (comprising a battery, bulb, and switch), believing voltage appears only once the switch is closed. This idea has been shown to be highly persistent even after instruction on introductory electricity and among more advanced learners [29,35–37].

4. Failure to understand electric circuits as a system

Students often lack a systemic perspective on electric circuits, leading to widespread student ideas such as “sequential reasoning” at secondary school and university levels. Sequential reasoning incorrectly attributes a temporal sequence to current flow through electric circuits. According to this student idea, in a single-loop closed circuit, the *order* of circuit elements such as resistors and light bulbs is believed to substantially affect their functioning, since the current is thought to pass sequentially through each component starting from the battery [36,38]. For example, in a closed circuit containing an ammeter positioned between two resistors, students frequently assume that the ammeter reading will change only if the resistance is altered at a point located “upstream” relative to the ammeter’s position in the direction of the current [14,29].

5. Resistance—failure to distinguish between parallel and series circuits

When asked about the total resistance of different resistor configurations, students frequently do not pay attention to the way the resistors are connected, instead primarily focusing on their quantity [15]. Such reasoning leads students, for example, to incorrectly conclude that two identical resistors connected in series have the same total resistance as two identical resistors connected in parallel. Similarly, students often erroneously assume that two equal resistors connected in parallel have a greater total resistance than a single resistor alone. While many students may employ known equations to calculate the total resistance and arrive at a mathematically correct result, the correct application of formulas does not necessarily indicate conceptual understanding.

6. Distinguishing between current and energy

In electric circuits, the amount of energy per second that is dissipated by resistors or light bulbs depends on both the

current flowing through the component and the voltage across it. However, in everyday language, at least in German, the term for electric current (“Strom”) is often used as a synonym for electrical energy [39]. As a result, if students cannot distinguish conceptually between these two quantities, their reasoning about energy flow in simple circuits can be expected to predominantly focus on current alone.

D. Assessment of student understanding and ideas

In multiple-choice test settings, single-tier items, which only require students to select one answer per question, may be fundamentally insufficient for a reliable diagnosis of specific student ideas, as the individual reasoning behind the answers given remains in the dark. A two-tier item design involves students selecting an answer to a main question (first tier) and then choosing one of several suggested reasons that best explains their initial answer (second tier). This approach allows for the identification of false-positive responses and reduces the likelihood of correct answers being selected by guessing. In a study on two-tier test instruments, Tamir [40] confirmed that the additional information provided in the second tier is valuable for diagnosing meaningful understanding. Consequently, modern test instruments in science, technology, engineering, and mathematics education research increasingly shift away from traditional multiple-choice formats in favor of at least a two-tier design [41–43]. However, some challenges associated with two-tier items include the use of test-taking strategies to select reasons rather than retrieval from memory, and the potential for students’ answer choices to be influenced by the second-tier options [44], which is particularly relevant in paper-and-pencil administrations. Some students were also found to consider incorrect answer options on the second tier, i.e., possible reasons, for the first time while completing the test [44].

In the following, we provide an overview of existing multiple-choice-type test instruments addressing the topic of simple electric circuits. We present and discuss their respective properties, strengths, and limitations with regard to (i) original language, (ii) main target group, (iii) scope of concepts covered, (iv) test and item design, and (v) psychometric quality, i.e., their ability to reliably assess students’ conceptual understanding and diagnose specific student ideas. Table II summarizes this analysis, forming the basis upon which we subsequently defined the objective for our own test development (see Sec. III). We limited the following analysis to test instruments addressing only the subject of simple, direct current electric circuits. More test instruments exist on the subject of electricity [45–47] in some cases also including items on dc circuits. However, these test instruments cover a wider scope of topics like electrostatics, electromagnetism, and electrodynamics, or ac circuits including elements like capacitors and inductors.

TABLE II. Overview of previously developed test instruments.

Name (original language)	No. of items	No. of tiers	Target group	Scope	Reliable assessment		
					Overall understanding	Individual student ideas	Individual concepts
Rhöneck [17] (GER)	12	1	High school	Current, voltage, resistance			
DIRECT [18] (ENG)	29	1	High school and university	Current, voltage, resistance, energy	•		
SECDT [49] (TUR)	12	3	High school	Current	•		
AECC [19] (GER)	23	1/2	High school	Current, resistance	•	•	
2T-SEC [20] (GER)	25	2	High school	Current, voltage, resistance	•		

Research on students' reasoning about electricity began in the 1980s. In the German-speaking context, the first major contribution toward the assessment of student ideas on simple electric circuits and the related key concepts was a study by Rhöneck [17]. This study describes several student ideas on current, voltage, and resistance that had previously been documented and categorized. In particular, Rhöneck provided a 12-item test which was administered to a sample of tenth-grade students across three European countries. The test items themselves vary considerably in format—some are multiple-choice, others single-choice, open-ended questions requiring numerical answers, or true-or-false questions. This can be understood against the backdrop that the primary aim of the study was to illustrate the prevalence of particular student misconceptions rather than to develop a comprehensive concept inventory. Moreover, each student idea was typically assessed by only one or two items. Therefore, the test does not meet contemporary psychometric standards required for reliably diagnosing specific student ideas [48]. However, with this work, Rhöneck significantly influenced subsequent research, and adaptations of certain items can still be found in current test instruments.

The English-language DIRECT test instrument [18], developed in the United States in 2004, represented a significant step forward. It comprises 29 single-tier, single-choice items on the concepts of current, voltage, resistance, and also energy transfer in simple electric circuits. It was evaluated statistically with a large sample of high school and university students, showing acceptable overall reliability. Student ideas underlying the item responses were identified retrospectively through interviews. In summary, DIRECT is a well-designed test instrument capable of assessing students' overall conceptual understanding of electric circuits with reasonable reliability. The 12 items of the “simple electric circuits diagnostic test” (SECDT) [49], originally developed in Turkish, include not only two tiers but also a third tier in which students are asked to self-report their confidence in their responses. Based on research by Hasan *et al.* [50], the developers suggest the

confidence rating as an easy means to distinguish between a “lack of knowledge” and “genuine misconceptions.” However, more in-depth research has shown that there is no clear-cut correlation between confidence in a wrong response and the underlying student idea being held consistently [27]. The test was designed based on the Turkish secondary school curriculum and was evaluated with a sample of 124 ninth-grade students from one particular district. While it primarily focuses on various student ideas related to electric current, it also contains one item on sequential reasoning. The reliability of the test was found to be low [49], and no advanced statistical analyses were conducted to evaluate the instrument's validity in assessing the student ideas as defined.

The first major German test instrument on simple electric circuits was developed by the AECC physics education research department at the University of Vienna in 2012. The 23-item test instrument, hereafter referred to as the AECCP test [19], includes both one-tier and two-tier items. It was psychometrically evaluated using a sample of grade 7 to 13 students across various school types in Austria and showed good overall reliability. The test was primarily designed to assess student ideas on the concepts of current and resistance, whereas other important concepts like voltage and energy flow in electric circuits are not addressed. However, each student idea is assessed by at least three items, which is an advantage compared to the instruments described above. Confirmatory factor analysis provided further evidence that reliable identification of certain student ideas is possible using this test. The “two-tier simple electric circuits” (2T-SEC) test instrument [20] can be viewed, in part, as a refinement and extension of the AECCP test, as all of its 25 items are two-tier and, in addition, they cover the concept of voltage. It is targeted at students at the lower secondary level and was evaluated using a large sample of German and Austrian students. Rasch analysis was used in its evaluation and showed good item fit statistics and high item reliability. Person reliability, comparable to Cronbach's alpha of classical test theory, however, was only marginally acceptable, and the test

proved to be too difficult for the sample used. No factor analyses were conducted to evaluate the instrument's suitability for identifying specific student ideas, as the test items were not primarily developed with a focus on student ideas but rather on physics concepts, e.g., current, resistance, and voltage.

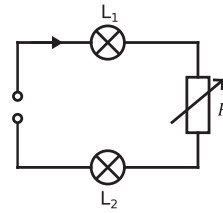
III. OBJECTIVE AND DEVELOPMENT PROCESS

The objectives guiding our test development are derived from the preceding analysis. We aim to develop a two-tier multiple-choice test instrument on simple electric circuits, specifically designed for undergraduate students majoring in physics and natural sciences. The instrument is intended to comprehensively cover all relevant key concepts (e.g., voltage and current) in a balanced manner. To validate this newly developed instrument, we employ both qualitative and quantitative methods. Our primary goal is to gather evidence supporting its validity, specifically focusing on its ability to reliably assess (i) students' overall conceptual understanding of simple electric circuits, (ii) students' propensity for specific student ideas, and (iii) students' understanding of individual key concepts [51].

The Basic Electricity Assessment of Students' Thinking (BEAST) was developed through a joint research project involving three universities from Germany and Austria. First, we identified six student ideas related to key concepts of electric circuits, known from prior research to be widespread among learners from secondary education through undergraduate physics students at the university level (see Sec. II). Second, we screened existing test instruments on simple electric circuits (see Sec. II D), selecting and adapting items suitable for assessing the key concepts and student ideas. Third, for each student idea, we initially developed four two-tier, single-choice test items, then selected three of those items, resulting in a final version of the instrument encompassing 18 items. For each item, only one combination of answer and reason is correct. Moreover, specific response combinations—typically one per item—reflect a specific student idea. It should be noted here that a direct association between students' manifest responses and presumed underlying student ideas aligns more closely with the framework theory perspective described in Sec. II B than with the knowledge in pieces view of students' conceptual reasoning. Nevertheless, without such coding, any quantitative evaluation of the psychometric properties of an instrument of this kind would be virtually impossible.

To ensure consistency, all item stems follow the same basic structure. Each describes a specific situation using text and one or more circuit diagrams, followed by the two-tier item format described above. On the second tier, we included an alternative in case participants could not find their own justification in any of the suggested options. Careful attention was given not only to ensuring that

In the circuit shown, two light bulbs and a variable resistor are connected to a battery.
The resistance R of the variable resistor is then increased.



How does this affect the brightness of the light bulbs L_1 and L_2 ?

- ☐ The brightness of both light bulbs decreases.
- ☐ The brightness of neither of the two light bulbs changes.
- ☐ The brightness of both light bulbs increases.
- ☐ Only the brightness of L_2 decreases.

How did you justify your answer above?

Important: Select the reason that best corresponds to your own thought process while answering the question.

- ☐ More current is required due to the increased resistance, which increases the current in the entire circuit.
- ☐ In the direction of current, only the light bulb that comes after the increased resistance is affected.
- ☐ The current in the entire circuit decreases due to the increased resistance.
- ☐ The brightness of the light bulb is only determined by the constant current supplied by the battery.
- ☐ I justified my answer in a completely different way.

FIG. 1. Example item (SEQ2) of the BEAST test instrument. The symbol on the left represents a general dc voltage source (as described in the participant instructions). The indication of positive and negative terminals has been deliberately omitted to avoid suggesting a specific direction of current.

the item descriptions and answer options are easily comprehensible by employing the simplest possible language but also to maintaining a consistent number of answer options across items addressing the same student idea. Figure 1 shows an example of an item focusing on the student idea of sequential reasoning. Where applicable, the sources of items adapted or inspired by existing test instruments are explicitly cited, reflecting, to the best of our knowledge, the earliest occurrence of an adapted item within any instrument. Item numbers in Table III correspond to their original enumeration in the referenced test instruments. In cases where single-tier items were adapted, we incorporated an additional tier by developing appropriate distractors.

As elaborated in Sec. II B, students' reasoning about individual electric circuits has been shown to be highly context sensitive. From a "p-prim" perspective on student thinking, this implies that although a student's response to one item may suggest a specific student idea, even minor changes in its superficial features could lead to a different conclusion. To account for this in our test instrument and to examine whether a particular idea is expressed consistently across similar situations, we included within each category several items that differed only slightly—for instance, by substituting a resistor with a light bulb in an otherwise identical circuit diagram.

All items were originally developed and evaluated in German and were translated into English for the purpose of this article. The full test instrument along with participant instructions can be found in Supplemental Material [52].

TABLE III. Overview of items and their sources.

Category/student idea	Item	Source
CC	CC1	Reference [18], item 8
	CC2	Reference [19], item 28 (not published in article)
	CC3	Own development
BSCC	BSCC1	Reference [20], item 5 (but parallel circuit)
	BSCC2	Similar to BSCC1
	BSCC3	Reference [20], item 4
VI	VI1	Reference [18], item 28 (adapted) (all three items)
	VI2	
	VI3	
SEQ	SEQ1	Reference [20], item 4
	SEQ2	Reference [18], item 26
	SEQ3	Reference [17], item 7
RES	RES1	Reference [18], items 5 and 14
	RES2	(all three items)
	RES3	
EN	EN1	Own development
	EN2	(all three items)
	EN3	

IV. QUALITATIVE VALIDATION—EXPERT SURVEY

Based on the items and proposed answer coding derived from our initial development outlined in Sec. III, we conducted an expert survey with two primary objectives. First, the comments from the expert survey informed revisions of item descriptions, circuit diagrams, wording, and content of answers as well as answer coding. Specifically, response patterns were excluded from representing certain student ideas if significant doubts were raised by the experts. Second, alongside insights gained from quantitative analyses (detailed in Sec. V), the expert survey was designed to further substantiate the content validity of our test.

A. Method

The expert survey comprised a document featuring all 18 two-tier items from our test, along with the initial instructions for participants. Five experts, actively engaged in the field of physics education research and possessing prior experience teaching physics either at the school or university level, participated in the expert survey. Importantly, these experts were not directly involved in the item development. Each expert received a spreadsheet detailing, for every two-tier item, the combination of answer and reason identified as correct by the item developers, as well as the combinations

associated with particular student ideas. Additionally, experts were provided with concise definitions of the student ideas addressed by the items. They were then asked to evaluate each item by: (i) indicating their agreement with the proposed coding of answers as either correct solutions or representations of specific student ideas, utilizing a scale from 1 (agree) through 2 (partially agree, with restrictions) to 3 (disagree); (ii) providing general remarks on the listed combinations of answers and reasons, including a brief explanation when their agreement with the proposed coding was incomplete; and (iii) offering general comments on the item as a whole.

B. Results

An overview of the experts' ratings on the answer coding of the individual items is shown in Table IV. In the following, answer codes given in the format (a, b), a and b correspond to the codes assigned to the first-tier and second-tier answers, respectively, as detailed in the test instrument available in Supplemental Material [52].

Based on the expert ratings regarding the correct answer codes, all answer codes were retained, as each received strong endorsement from the majority of experts, with no answer being fully rejected by any reviewer. In the following, we provide an overview of some comments made by the experts for those items where judgments were not unanimously in favor (rating 5-0-0 in Table IV). For example, the expert who provided conditional agreement ("agreed with restrictions") on all three CC items, noted that the proposed reason does not exclude the view that other appliances or light bulbs might exist where current is indeed consumed. Regarding items BSCC1 and BSCC2, one expert noted that the proposed reason, "There is an additional current through the second bulb/resistor," might erroneously be interpreted as the bulb itself "causing" the additional current. However, to ensure consistency in length and complexity across all distractors, we opted to retain the reason initially proposed. Concerning the coding of student ideas, we decided to exclude the answer combinations (1, 3) of item VI2 and (4, 1) of item VI3 from representing the student idea VI (failure to distinguish between current and voltage) due to mixed expert feedback. However, these answers were retained as distractors within the instrument. Answer (1, 3) for item VI2 was primarily criticized by the experts for not aligning with the interpretation of answer (2, 1), which received unanimous agreement, but rather expressing the opposite ("only voltage when there is no current"). While answer (1, 3) was intended to represent the student idea that voltage "flows together with current" (i.e., a confusion of the two concepts), the experts noted that this is not the same as the student idea no voltage without a current. We agreed with this assessment and decided to focus the evaluation on the latter; the same rationale applies to answer (4, 1) for item VI3.

TABLE IV. Expert ratings on the item coding. *The answer codes are given in the format “first-tier answer, second-tier answer.” Answers in brackets were removed from further evaluation based on expert feedback. **Expert ratings are given as fully agree—agree with restrictions—do not agree.

Category/student idea	Item	Answer code* (correct)	Expert rating** (correct)	Answer code(s)* (student idea)	Expert rating** (student idea)
CC	CC1	3, 2	4-1-0	1, 1 2, 3	5-0-0 5-0-0
	CC2	1, 1	4-1-0	2, 2 3, 3	5-0-0 5-0-0
	CC3	1, 3	4-1-0	2, 1 3, 2	5-0-0 5-0-0
BSCC	BSCC1	1, 2	4-1-0	3, 1	5-0-0
	BSCC2	1, 2	4-1-0	3, 1	5-0-0
	BSCC3	2, 2	5-0-0	3, 1	5-0-0
VI	VI1	2, 1	4-1-0	1, 2	5-0-0
	VI2	4, 2	5-0-0	2, 1 (1, 3)	5-0-0 1-4-0
	VI3	1, 3	5-0-0	3, 2 (4, 1)	4-1-0 2-3-0
SEQ	SEQ1	2, 1	5-0-0	3, 3	5-0-0
	SEQ2	4, 1	4-1-0	3, 3	5-0-0
	SEQ3	3, 4	4-1-0	2, 2	5-0-0
RES	RES1	1, 3	4-1-0	2, 1 3, 4	5-0-0 4-0-1
	RES2	3, 3	4-1-0	1, 1	5-0-0
	RES3	1, 2	4-1-0	2, 1	5-0-0
EN	EN1	1, 2	5-0-0	2, 1	5-0-0
	EN2	2, 2	5-0-0	3, 1	5-0-0
	EN3	1, 2	5-0-0	2, 1	5-0-0

V. QUANTITATIVE VALIDATION

To evaluate the 18 two-tier items and the test instrument as a whole, we conducted a number of quantitative analyses. Where applicable, we followed the validation framework proposed by Jorion *et al.* [51] and interpreted the results according to their “categorical judgement scheme.” The analyses include methods of classical test theory as well as structural analyses like exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). An overview of the quantitative methods used for validating the BEAST test instrument is provided in Table V.

A. Test administration method and sample

The instrument was administered to a total of $N = 176$ undergraduate students at six German universities. In the

summer semester of 2023, a total of $N = 93$ students participated, and $N = 83$ students participated in the summer semester of 2024. All 18 items were administered using an online survey platform (SoSci Survey). Participants who did not complete all items or who responded too quickly—resulting in a “relative speed index” (RSI) of 2.0 or higher [54] as calculated by the survey platform—were excluded from the analysis. The following analyses are based on the remaining $N = 160$ participants (85 female, 70 male, 5 other/not specified, median age 21 years). All basic sociodemographic data—age, gender, university, course of study, and semester—were collected *after* participants had completed the electricity section of the questionnaire. In addition, we asked whether they had previously attended university-level courses (lectures or seminars) covering electricity,

TABLE V. Summary of quantitative analyses.

Analysis	Measures
Distractor analysis	Percentage of students selecting each answer option
Classical test theory	
Item statistics	Item difficulty Item discrimination
Total score reliability	Cronbach's alpha of total score Cronbach's Alpha-with-item-deleted
Structural analyses	
Exploratory factor analysis	Fit of the model to the data
Confirmatory factor analysis (bifactor models)	Item loading Comparative fit index Root-mean-square error of approximation Total reliability ω [53] Reliability ω_h of subscales [53]

which none of them had. All the symbols and the general test procedure were explained on an introductory page. Students were instructed to complete the test individually at home, without consulting external sources such as the internet, course materials, or other individuals. The time available to complete the test was not restricted, with a median completion time of 21 min for all 18 items.

The online format helped minimize distractions by presenting the items to participants one at a time and by partially randomizing their order. It also allowed us to address certain limitations of traditional paper-and-pencil testing with two-tier items as discussed in Secs. IID and IIB—particularly by preventing undesired strategies, such as selecting a “matching” answer after

viewing the reasoning options. To achieve this, the item tiers were presented sequentially: for each item, the first-tier question had to be answered before the second-tier options (possible reasons) were displayed; once submitted, first-tier responses could no longer be changed.

B. Distractor analysis

For each of the 18 items, we analyzed the frequency with which participants selected the correct response, defined as choosing the correct answer on the first tier along with the correct reason on the second tier. Similarly, we examined how often participants selected distractors, i.e., specific combinations of answer and reason that reflect a specific student idea. Figure 2 shows the percentages for both correct responses (white bars) and distractor selections across all 18 items, categorized by student ideas (gray bars).

The distractors corresponding to the various student ideas were selected with notable frequency, ranging from 16.2% (RES3) to 50.6% (BSCC1, UI1). The remaining proportion up to 100% comprises responses that were neither correct nor aligned with the intended student idea. This may result from random guessing, false positives (selecting the right answer but for the wrong reason), or false negatives (selecting an incorrect answer but without choosing the reason associated with the student idea). Part of this proportion is also due to participants choosing the “none” option on the second item tier, as we did not force a choice there. On average, this option was chosen in 13.2% of all cases across both administrations. This number is not surprising as there are obviously ways of thinking about the given problems not represented in the reasoning options, e.g., trying to use known equations instead of reasoning on a conceptual level. Nevertheless, even the lowest distractor selection rates far exceed chance level, indicating that the items are suitable for assessing the presence of specific student ideas.

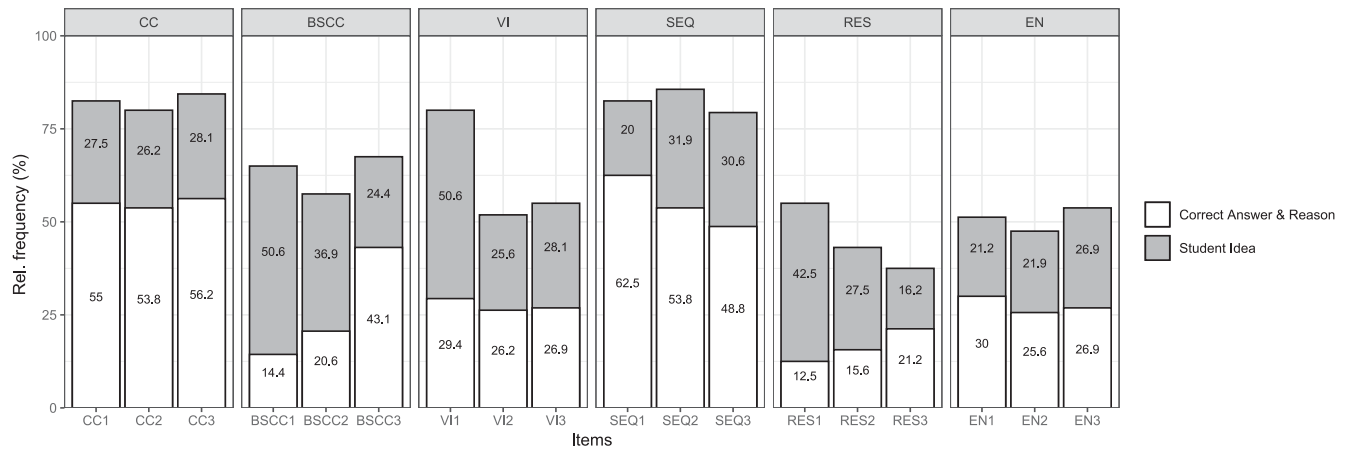


FIG. 2. Percentage of students selecting either the correct answer combination or the answer combination indicative of a student idea for all 18 items.

C. Classical test theory

1. Item difficulty and discrimination

Item difficulty, defined as the proportion of students who answered an item correctly, was calculated for each of the 18 items and is presented in Fig. 2. The percentage of correct responses ranged from 12.5% (RES1, indicating a difficult item) up to 62.5% (SEQ1, indicating a relatively easy item). Items from categories like CC and SEQ tended to be easier, whereas items from categories like VI or RES were more difficult. Overall, the range of item difficulties can be considered “good” according to established guidelines [51].

To assess item discrimination, the point-biserial correlation between each item score and the total test score (excluding the item in question) was calculated. This metric indicates how well an individual item distinguishes between high- and low-performing students. Item discrimination values should be positive, with higher values reflecting stronger alignment with overall test performance [55]. In our analysis, discrimination values range from 0.38 to 0.69, which is considered “excellent” [51,56]. Figure 3 summarizes the item statistics in a 2D plot.

2. Total score reliability (Cronbach’s alpha)

Cronbach’s alpha for the total score should be interpreted cautiously, as the BEAST test instrument was constructed based on multiple different physical concepts and corresponding student ideas. This likely results in a multidimensional structure, whereas Cronbach’s alpha assumes unidimensionality—that is, that all items reflect the same

underlying construct. Moreover, for Cronbach’s alpha to serve as a direct measure of reliability, the items must meet additional restrictive assumptions regarding their measurement properties, such as tau equivalence [53]. This will be expanded on in Sec. VD, where we present structural analyses of the test and report reliability scores for the subscales identified. With these limitations in mind, we still report Cronbach’s alpha of the test, which is 0.903. The “alpha-if-item-deleted” values range between 0.893 and 0.902, indicating that the overall alpha does not change substantially regardless of which item is hypothetically removed. While this suggests a degree of redundancy among items—an expected result given the conceptual overlap—it also indicates a high level of internal consistency. Taken together, and with appropriate caution, the total Cronbach’s alpha can be interpreted as excellent [51].

D. Structural analyses of the test

Structural analyses were employed as a further validation method to assess whether item responses could plausibly be attributed to the latent traits upon which the items were constructed. This analysis was conducted from two complementary perspectives. One is focused on conceptual understanding, as reflected in the correct selection of both answer and reason. Here, the underlying latent traits were assumed to represent the understanding of the basic concepts of electricity like current, resistance and voltage. The second perspective is centered on student ideas, operationalized through the selection of certain combinations of answer and reason indicative of student ideas. The

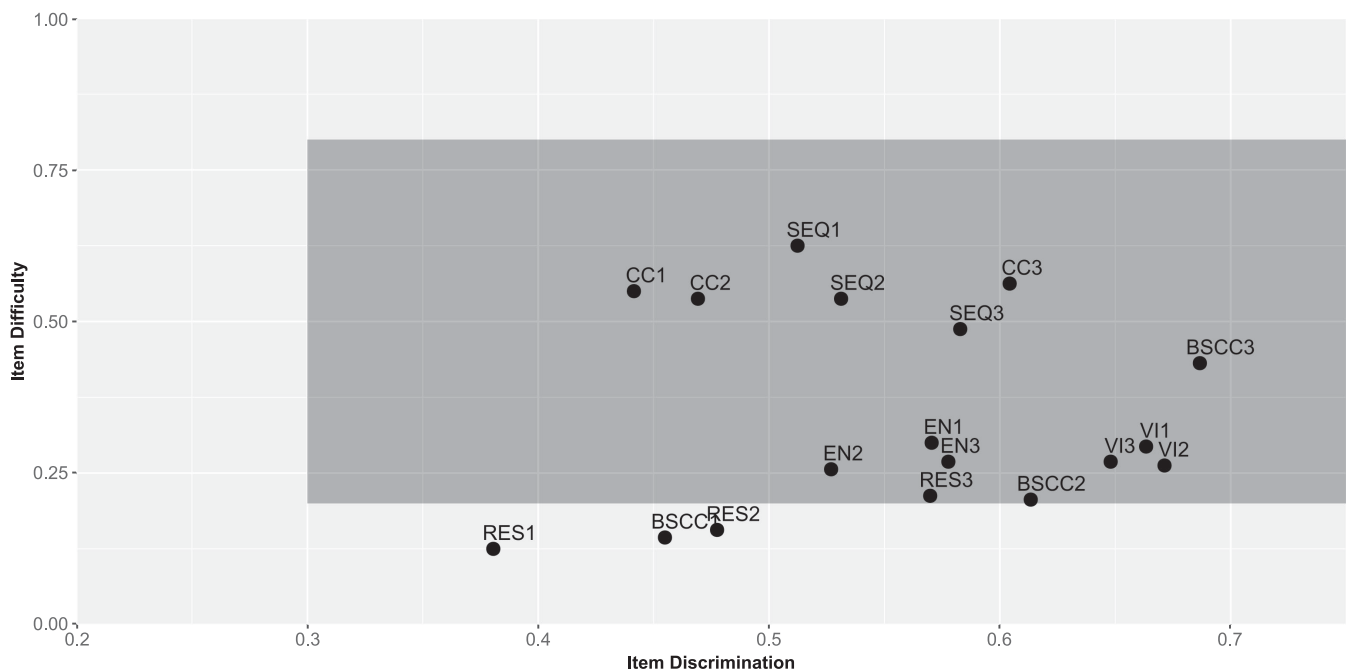


FIG. 3. Item statistics for all 18 items included in the test. The area considered ideal is highlighted in dark gray.

alignment of the item groups with these concepts and student ideas is detailed in Sec. II C. For both perspectives, the same analytic strategy—combining exploratory and confirmatory approaches—was applied. As a first step, a parallel analysis was conducted in order to estimate a plausible range of values for the number of underlying factors. Based on these results, an exploratory factor analysis (EFA) was performed, specifying the number of factors suggested by the parallel analysis. The resulting factor structures were then compared with the expected structure, which was based on the theoretical assumption that each group of three items should reflect a single latent trait (i.e., one for each student idea or concept). Both parallel analysis and EFA were conducted using the PSYCH package in R.

Using these results as a starting point, we used the LAVAAN package of R to conduct confirmatory factor analyses with various structural equation models that should fit the data from a theoretical point of view. Model quality was evaluated using common fit indices, including root-mean-square error of approximation (RMSEA), Tucker-Lewis index (TLI), comparative fit index (CFI), and factor loadings.

1. Structural analysis based on student ideas

A dichotomous model was applied to determine the presence (1) or absence (0) of a student idea, based on whether the corresponding combination of answer and reason was selected. Parallel analysis (factor method: minimum residual solution, eigenvalues: principal axis factor analysis) reveals the presence of a dominant factor accounting for a substantial portion of the total variance. This is indicated by the large eigenvalue of the first principal factor in the scree plot (Fig. 4) and the much lower eigenvalues of the other factors. Moreover, the parallel analysis suggests that a total number of five or six (progressively weaker) factors may be reasonable. This inference is based on comparisons between the eigenvalues of the actual data and those of simulated data. Therefore,

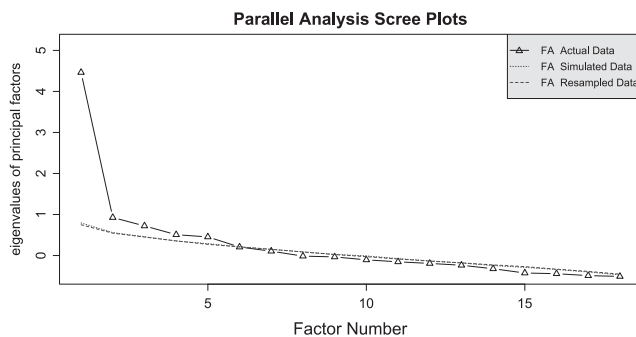


FIG. 4. Scree plot displaying the eigenvalues of the principal factors based on student ideas.

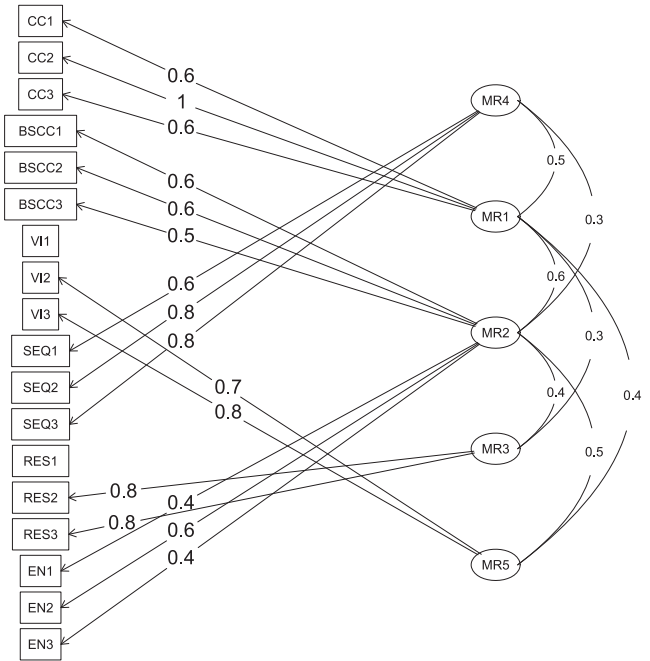


FIG. 5. Factor solution showing the factors MR1–MR5 and their loadings for all 18 items, as determined by EFA. Loadings below 0.3 are not displayed.

we performed an EFA specifying both five and six factors, using oblique rotation methods (“oblimin” and “promax”), which is theoretically justified given that student conceptions of electric circuits are likely to be interrelated [39]. A factor solution with five factors resulted in the best, but only marginally acceptable model fit (RMSEA = 0.071 with a 95% confidence interval of [0.052; 0.091], TLI = 0.84). The extracted five-factor structure, derived using promax rotation, is shown in Fig. 5.

Most of these factors can be meaningfully interpreted and align with the intended constructs assessed by the items. Factor MR4 is strongly associated with all three SEQ items and therefore seems to represent sequential reasoning. All three CC items show loadings on MR1, so this factor likely corresponds to the student idea of current consumption. MR2 is associated with the idea of sequential reasoning as it is linked to the three SEQ items. MR3 and MR4 show the highest factor loadings on two out of three items of the categories RES and VI, respectively. However, some deviations from the expected structure were observed. For example, RES1 and VI1 did not load above 0.3 on any factor, and the items from the BSCC and EN categories do not form distinct factors, instead loading together on MR5 along with the other BSCC items. These irregularities notwithstanding, all factors in the solution were found to be strongly correlated. This is theoretically plausible as student ideas such as “batteries supply constant current” (BSCC), “current is consumed” (CC), and “sequential reasoning” (SEQ) are often used in combination when reasoning about electric circuits [39].

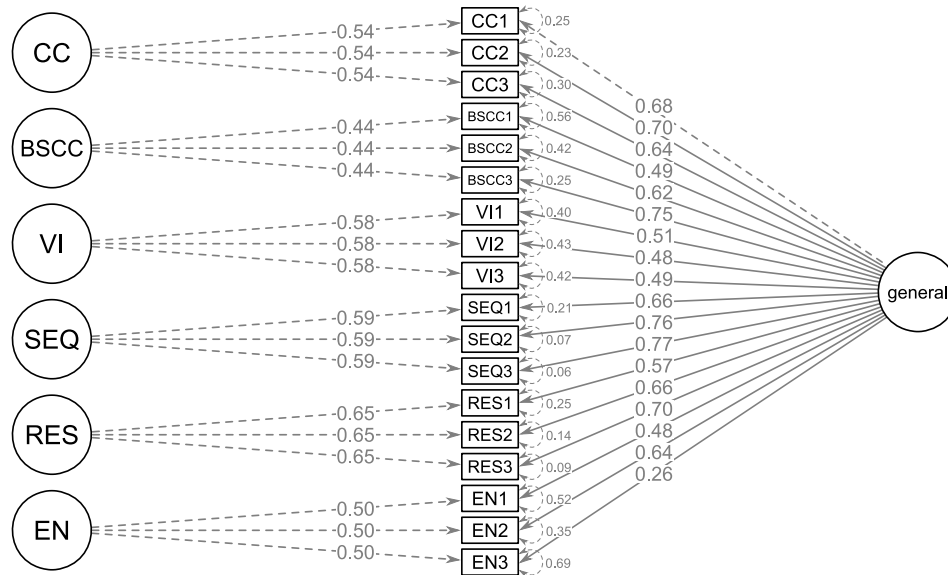


FIG. 6. Factor solution for the bifactor model, featuring one general factor and six orthogonal group factors representing the individual student ideas.

To further assess the underlying structure, a confirmatory factor analysis (CFA) using the `LAVAN` package in R was conducted. As a baseline, we tested a simple unidimensional model; a good model fit in this case would indicate that all items measure a single underlying construct. The model fit for the unidimensional model was only marginally acceptable (RMSEA = 0.094 with a 95% confidence interval of [0.081; 0.107], TLI = 0.889, CFI = 0.902). As a next step, we tested a six-factor model in which each factor was hypothesized to represent a specific student idea and was therefore aligned with the corresponding group of three items. This model showed substantially improved fit indices (RMSEA = 0.051 with a 95% confidence interval of [0.033; 0.067], TLI = 0.967, CFI = 0.972), although the factors were strongly correlated, with standardized covariances ranging from 0.40 to 0.70. As anticipated from the previous analyses, a six-factor model with uncorrelated (orthogonal) factors did not yield an acceptable fit.

Given these findings, which suggest the presence of one dominant general factor alongside several weaker specific factors (likely representing student ideas), we tested a bifactor model. This model included six orthogonal group factors and an additional general factor for all items. To maintain parsimony and account for the limited sample size, we constrained the factor loadings to be equal within each group factor. The bifactor model also showed a good fit (RMSEA = 0.0486 with a 95% confidence interval of [0.026; 0.063], TLI = 0.973, CFI = 0.977). The resulting factor solution, including residual variances, is shown in Fig. 6. All factor loadings are highly significant.

From a framework theory perspective on student thinking, the individual “group factors” or “latent traits” in this

structural equation model can be interpreted as deeply rooted cognitive structures that represent distinct, theory-like ideas as described in Sec. II B. For most items, the factor loadings of the general factor exceed those of the individual factors. While this result cannot be fully explored within the scope of the current data, the general factor may (cautiously) be interpreted as a general tendency to hold common student ideas about electric circuits and concepts of electricity. We expand on possible ways of interpreting this result in Sec. VI C.

The total reliability of the bifactor model was estimated at $\omega = 0.930$ by the “`compRelSEM`” function from the `SEMTtools` package in R, indicating that a substantial portion of the total observed variance is explained by the model’s factors. The hierarchical reliability scores ω_h of the subscales are reported in Table VI, reflecting the proportion of item response variance explained by each specific factor.

TABLE VI. Hierarchical reliability coefficients of the subscales and the general factor in the bifactor model based on student ideas.

Subscale/factor	ω_h
General	0.785
CC	0.273
BSCC	0.202
VI	0.362
SEQ	0.273
RES	0.425
EN	0.278

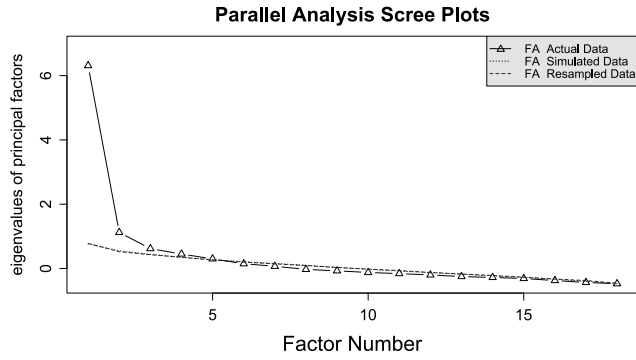


FIG. 7. Scree plot displaying the eigenvalues of the principal factors from the analysis based on conceptual understanding.

2. Structural analysis based on conceptual understanding

Similar to the previous analysis, a dichotomous scoring model was applied, whereby an item was scored as correct (1) if both the correct answer and the correct reason were given; all other responses were scored as incorrect (0). This allows for a structural analysis based on participants' conceptual understanding. However, these results should be interpreted with caution, as the items were primarily developed based on student ideas rather than the understanding of concepts directly. In other words, avoiding certain student ideas by selecting the correct answer does not necessarily indicate a full understanding of the associated physical concept. For example, a student may know that current is not consumed and that a battery does not supply a constant current, while still lacking an adequate conceptual understanding of current itself.

TABLE VII. Hierarchical reliability coefficients of the subscales and the general factor in the bifactor model based on conceptual understanding.

Subscale/factor	ω_h
General	0.866
Current	0.312
Battery	0.070
Voltage	0.118
System	0.295
Resistance	0.233
Energy	0.193

First, we conducted a parallel analysis. The corresponding scree plot (Fig. 7) reveals one dominant factor alongside several smaller factors influencing item responses. Based on this, a total number of two to four factors seems reasonable. We subsequently performed an EFA using different oblique rotation methods. The four-factor model showed the best, if only marginally acceptable, model fit (RMSEA = 0.080 with a 95% confidence interval of [0.064; 0.098], TLI = 0.858). As expected, the factors were highly correlated, with correlation coefficients ranging from 0.29 to 0.63. As a result, none of the extracted factors in the solutions using either rotation method could be clearly aligned with the physical concepts associated with the item groups. Therefore, we proceeded with a CFA, fitting a bifactor model to the data. The model structure closely mirrored that of the previous analysis, comprising one general factor and six specific group factors representing the conceptual categories most closely related to each item

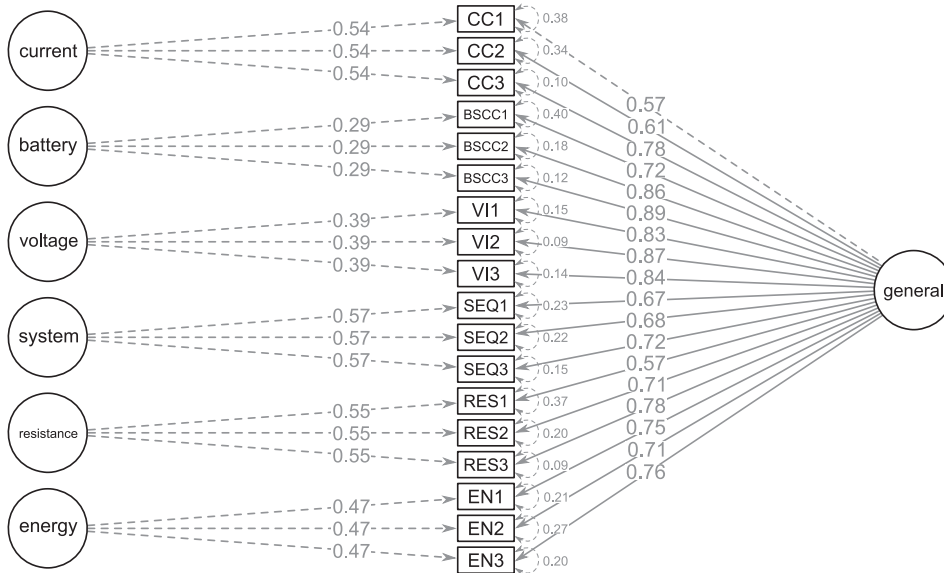


FIG. 8. Factor solution for the bifactor model, featuring one general factor and six orthogonal group factors modeling the understanding of individual concepts.

set (see Sec. II C). As before, factor loadings were constrained to be equal within each group factor. The bifactor model demonstrated a good fit (RMSEA = 0.044 with a 95% confidence interval of [0.022; 0.061], TLI = 0.982, CFI = 0.985). The resulting factor structure is shown in Fig. 8. Again, from a framework theory perspective on student thinking, the individual group factors can be understood as cognitive traits reflecting a participant's way of making sense of the various concepts.

Across all items, most of the total variance is explained by the general factor, suggesting that the test predominantly measures a single overarching construct. Nevertheless, a smaller, yet highly significant portion of variance is explained by the specific factors. Some items show a relatively high residual variance, which may partially reflect limitations of the simplified model. Based on this bifactor model, the total reliability was estimated at $\omega = 0.965$ using the compRelSEM function. The hierarchical reliability scores ω_h of the subscales are reported in Table VII.

VI. DISCUSSION, LIMITATIONS, AND OUTLOOK

A. Discussion

In this article, we presented the development and statistical evaluation of a test instrument on simple electric circuits designed to assess undergraduate students' conceptual understanding and common student ideas on simple electric circuits. The content validity and relevance of the test items were ensured through an extensive literature review on student ideas, analysis of existing test instruments, the collective expertise of the development team, and by conducting a comprehensive expert survey.

Based on quantitative analyses, there is substantial evidence supporting the claim that the BEAST test instrument meets the three development objectives outlined in Sec. III. First, multiple findings support that the objective is met, namely that (i) the test instrument reliably assesses students' overall conceptual understanding. Item difficulty and discrimination analysis (Sec. V C 1) indicate that the items are well matched to the ability levels of the target population and contribute meaningfully to differentiating between higher and lower levels of understanding. The value of Cronbach's alpha for the total score (0.903) indicates a high internal consistency, although this should be interpreted cautiously due to the multidimensional nature of the instrument (Sec. V C 2). Furthermore, confirmatory factor analysis using a bifactor model based on conceptual understanding yields a high total reliability estimate of $\omega = 0.965$. The presence of a strong general factor explaining a large portion of the observed variance ($\omega_h = 0.866$) further confirms that the test captures an overarching conceptual understanding of electric circuits (Sec. V D 2).

Second, there are several lines of evidence that the objective is met, namely that (ii) the test instrument can assess students' propensity for specific student ideas. First, the expert survey revealed strong agreement that, for the majority of items, the distractors in the two-tier design are representative of the intended student ideas (Sec. IV B). Where this was initially not the case, items or coding schemes were revised accordingly. Second, distractor analysis shows that student ideas were frequently selected by participants (Sec. V B). Third, exploratory factor analysis using correlated factors generally reproduces the expected factor structure. Fourth, confirmatory factor analysis using a bifactor model reflecting the hypothesized student ideas as individual factors shows a good model fit and significant loadings on latent factors modeling the various student ideas (Sec. V D 1).

Finally, there is also support that the objective is met, namely that (iii) the instrument can assess the understanding of individual key concepts. First, the expert survey confirmed that the proposed correct combinations of answers were widely accepted as valid indicators of correct conceptual understanding (Sec. IV B). Second, confirmatory factor analysis based on conceptual understanding using a bifactor model shows a good model fit, and the latent factors representing fundamental electricity concepts (e.g., voltage or current) exhibit highly significant loadings (Sec. V D 2).

The publication of the BEAST test instrument will enable educators to use it as a tool for formative assessment, to monitor student progress and provide targeted support based on the student ideas diagnosed to be most prevalent. Additional applications include the summative assessment of students' conceptual understanding of electric circuits or the evaluation of instructional interventions, for example, using a pre-post study design. As the test instrument can be implemented as an online survey, it can be easily administered to large student groups, and automated feedback systems can be readily integrated.

B. Limitations

Despite the strengths of the BEAST test instrument discussed, several limitations should be acknowledged.

First, the test instrument does not cover the full spectrum of student ideas on simple electric circuits. Given the constraints on test length and the requirement that each subscale include at least three items, we prioritized student ideas that are both well documented in the literature and closely tied to core concepts such as current, voltage, and resistance. Consequently, less prevalent or more nuanced student ideas could not be included.

Second, it is important to note that the primary goal of the test development was to assess specific student ideas. While correctly answering an item may indicate that a student does not hold a specific student idea, it does not necessarily imply a complete or accurate understanding of

the underlying physical concept. For example, a student might correctly reject the idea that current is consumed or that batteries supply a constant current and still harbor other conceptual difficulties about the nature of current or the functioning of batteries. Therefore, analysis focusing on conceptual understanding should be interpreted with caution. In particular, the factors identified in structural analysis should be viewed as representing the concepts most closely associated with the student idea of the respective item category, rather than comprehensive indicators of conceptual mastery.

Third, regarding our sample, it should be noted that the number of participants ($N = 160$) used to estimate our model parameters is relatively small compared to common recommendations for structural equation modeling [57] (pp. 14–16). To address this limitation, we employed models that were as parsimonious as possible. Although the data fit well with the bifactor models presented in this article, and we are therefore confident in the robustness of our interpretations, the findings should nevertheless be interpreted with appropriate caution, given the sample size. Also, with regard to our sample, although we asked participants whether they had previously attended university-level courses covering electricity, their familiarity with dc circuits may nonetheless vary due to other individual factors that we did not assess in detail, such as prior degrees, completed vocational training, or personal interests. However, we have no indication that our sample was in any way atypical compared to any other sample of first-year undergraduate students.

C. Outlook

In both bifactor models, we found that the general factor showed higher factor loadings than the factors representing individual student ideas or concepts for nearly all items. Although the models fit well to the data and specific factors were clearly identifiable; this result implies that a substantial portion of the total observed variance results from the general factor. This is further reflected by the high ω_h value for the general factor and the comparatively low values for the subscales associated with student ideas or concepts. The relatively low reliability of the subscales may reflect a high context dependency of student ideas in general and in the domain of electric circuits in particular (see Sec. II B).

In the following, we offer some reflections on possible explanations for this pattern, which may serve as a basis for future research. From a theoretical point of view, in a bifactor model, the general factor reflects the unidimensional aspect of a construct. Although the test instrument was explicitly designed to address distinct student ideas—implying a multidimensional structure—a certain degree of unidimensionality is expected, simply because all items share a common content domain: electric circuits. In the bifactor model based on student ideas (Sec. V D 1), the

general factor might be interpreted as a “general susceptibility to holding student ideas about electric circuits.” Accordingly, in the model based on conceptual understanding (Sec. V D 2), it may reflect a “general understanding of electric circuits.” While such broad labels are helpful for orientation, they lack specificity. It remains an open question what underlying traits this general factor actually represents. For example, the general factor may reflect individual differences in familiarity with or interest in electric circuits, which could influence overall performance. Alternatively, the general factor may incorporate broader competencies not specific to electric circuits, such as test-taking skills or reading comprehension, or broader cognitive abilities like general scientific literacy and reasoning skills. Controlling for all these variables would require additional instruments and more complex structural equation models—efforts that would demand larger sample sizes. Nonetheless, it would be worthwhile to investigate why, despite the targeted design of the items, student ideas appear to exert only a limited influence on response patterns. Future studies might also benefit from qualitative approaches, such as conducting interviews or think-aloud studies with students, which would provide deeper insight into how students interpret and respond to the items and allow for a more thorough evaluation of item validity, enriching the existing insights gained from the expert surveys.

Another interesting research direction involves examining the relationship between the consistency with which learners exhibit particular student ideas and their self-reported certainty, following the work of Hull *et al.* [27]. The BEAST test instrument presented in this paper is well suited for this purpose, as each student idea is assessed by a set of three items, allowing consistency to be quantitatively measured. In a future study, a third tier could be added to each item, asking students to rate how confident they are in the correctness of their answer. Preliminary findings from the present study suggest that student ideas were generally expressed inconsistently across the three items of every category. Investigating whether this inconsistency correlates with students’ confidence ratings could yield valuable insights into the metacognitive dimensions of their reasoning and could specifically contribute to gathering more evidence in the ongoing theoretical knowledge in pieces vs framework theory discussion.

ACKNOWLEDGMENTS

We acknowledge the support by the Open Access Publication Fund of the University of Tübingen.

DATA AVAILABILITY

The data that support the findings of this article are openly available [58].

- [1] Ministerium für Kultus, Jugend und Sport Baden-Württemberg, Bildungsplan des Gymnasiums 2016—Physik, https://www.bildungsplaene-bw.de/site/bildungsplan/get/documents/lbw/export-pdf/depot-pdf/ALLG/BP2016BW_ALLG_GYM_PH.pdf (2016) [accessed September 23, 2025].
- [2] Bundesministerium für Bildung Österreich, Lehrpläne—allgemeinbildende höhere Schulen, <https://www.ris.bka.gv.at/Dokumente/Bundesnormen/NOR40264237/NOR40264237.pdf> (2024) [accessed September 23, 2025].
- [3] Department for Education, Science programmes of study: Key stages 1 and 2—National curriculum in England, https://assets.publishing.service.gov.uk/media/5a806ebd40f0b62305b8b1fa/PRIMARY_national_curriculum_-_Science.pdf (2013) [accessed September 23, 2025].
- [4] Department for Education, Science programmes of study: Key stage 3—National curriculum in England, https://assets.publishing.service.gov.uk/media/5a7d563de5274a2af0ae2ffa/SECONDARY_national_curriculum_-_Science_220714.pdf (2013) [accessed September 23, 2025].
- [5] Ministère de l'Éducation nationale, de la Jeunesse et des Sports, Le bulletin officiel de l'Éducation nationale, de la jeunesse et des sports—annexe 3: Programme d'enseignement du cycle des approfondissements (cycle 4), https://cache.media.education.gouv.fr/file/31/89/1/ense1714_annexe3_1312891.pdf (2020) [accessed September 23, 2025].
- [6] Ministère de l'Éducation nationale, de la Jeunesse et des Sports, Curriculum for compulsory school, preschool class and school-age educare—Lgr22, <https://www.skolverket.se/getFile?file=13128> (2024) [accessed September 23, 2025].
- [7] Next generation science standards—MS-PS evidence statements, https://www.nextgenscience.org/sites/default/files/evidence_statement/black_white/MS%20PS%20Evidence%20Statements%20Dec%202020%20asterisks.pdf (2015) [accessed September 23, 2025].
- [8] Landesinstitut für Schulentwicklung Baden-Württemberg, Bildungsplan 2016—Beispielcurriculum für das Fach Physik—Klassen 7/8, Beispiel 1, https://www.schule-bw.de/service-und-tools/bildungsplaene/allgemein-bildende-schulen/bildungsplan-2016/beispielcurricula/gymnasium/BP2016BW_ALLG_GYM_PH_BC_7-8_BSP_1.pdf (2016) [accessed March 20, 2025].
- [9] H. Muckenfuß and A. Walz, *Neue Wege Im Elektrizitätsunterricht: Vom Tun über Die Vorstellung Zum Begriff* (Aulis-Verlag Deubner, Köln, 1992).
- [10] M. S. Steinberg, *Electricity Visualized — The CASTLE project. Student Manual* (Pasco Scientific, Rosevill, CA, 2004).
- [11] J.-P. Burde and T. Wilhelm, Teaching electric circuits with a focus on potential differences, *Phys. Rev. Phys. Educ. Res.* **16**, 24 (2020).
- [12] P. V. Engelhardt, K. E. Gray, and N. S. Rebello, How many students does it take before we see the light?, *Phys. Teach.* **42**, 216 (2004).
- [13] 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).
- [14] R. Cohen, B. Eylon, and U. Ganiel, Potential difference and current in simple electric circuits: A study of students' concepts, *Am. J. Phys.* **51**, 407 (1983).
- [15] 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).
- [16] M. K. R. Stetzer, P. van Kampen, P. S. Shaffer, and L. C. McDermott, New insights into student understanding of complete circuits and the conservation of current, *Am. J. Phys.* **81**, 134 (2013).
- [17] C. v. Rhöneck, Vorstellungen vom elektrischen Stromkreis und zu den Begriffen Strom, Spannung und Widerstand, *Naturwiss. Unterr. Phys.* **34**, 10 (1986), <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?Fid=2139095>.
- [18] P. V. Engelhardt and R. J. Beichner, Students' understanding of direct current resistive electrical circuits, *Am. J. Phys.* **72**, 98 (2004).
- [19] H. Urban-Woldron and M. Hopf, Entwicklung eines Testinstruments zum Verständnis in der Elektrizitätslehre, *Z. Didakt. Naturwiss.* **18**, 201 (2012).
- [20] L. Ivanjek, L. Morris, T. Schubatzky, M. Hopf, J.-P. Burde, C. Haagen-Schützenhöfer, L. Dopatka, V. Spatz, and T. Wilhelm, Development of a two-tier instrument on simple electric circuits, *Phys. Rev. Phys. Educ. Res.* **17**, 020123 (2021).
- [21] R. Duit, Bibliography: Students' and teachers' conceptions and science education, <http://archiv.ipn.uni-kiel.de/stcse/> (2009).
- [22] R. Driver, Pupils' alternative frameworks in science, *Eur. J. Sci. Educ.* **3**, 93 (1981).
- [23] A. A. diSessa, Knowledge in pieces, in *Constructivism in the Computer Age*, The Jean Piaget Symposium Series (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1988), pp. 49–70.
- [24] E. Van den Berg and L. Sandura, Student ideas on the velocity of light, *Aust. Sci. Teach. J.* **36**, 72 (1990).
- [25] S. Vosniadou and I. Skopeliti, Conceptual change from the framework theory side of the fence, *Sci. Educ.* **23**, 1427 (2014).
- [26] A. A. diSessa, Toward an epistemology of physics, *Cognit. Instr.* **10**, 105 (1993).
- [27] M. M. Hull, A. Jansky, and M. Hopf, Does confidence in a wrong answer imply a misconception?, *Phys. Rev. Phys. Educ. Res.* **18**, 020108 (2022).
- [28] R. Duit, E. Kircher, R. Girwidz, and P. Häußler, Alltagsvorstellungen und Physiklernen, in *Physikdidaktik: Theorie und Praxis* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 581–606, [10.1007/978-3-540-34091-1_19](https://doi.org/10.1007/978-3-540-34091-1_19).
- [29] D. M. Shipstone, A study of children's understanding of electricity in simple dc circuits, *Eur. J. Sci. Educ.* **6**, 185 (1984).
- [30] R. Osborne, Towards modifying children's ideas about electric current, *Res. Sci. Technol. Educ.* **1**, 73 (1983).
- [31] D. M. Shipstone, C. V. Rhöneck, W. Jung, C. Kärrqvist, J.-J. Dupin, S. Johsua, and P. Licht, A study of students' understanding of electricity in five European countries, *Int. J. Sci. Educ.* **10**, 303 (1988).

- [32] L. McDermott and E. Van Zee, Identifying and addressing student difficulties with electric circuits, in *Aspects of Understanding Electricity* (IPN, Kiel, Germany, 1985), Vol. 39.
- [33] J.-J. Dupin and S. Johsua, Conceptions of french pupils concerning electric circuits: Structure and evolution, *J. Res. Sci. Teach.* **24**, 791 (1987).
- [34] C. Rhoeneck and B. Voelker, Vorstellungen vom Stromkreis und ihr Einfluss auf den Lernprozess, *Der Physikunterricht*. **18**, 4 (1984), <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?Fid=2096585>.
- [35] C. Rhöneck, Schüleräußerungen zum Begriff der elektrischen Spannung beim Erklären der Meßwerte am Schalter, *Naturwiss. Unterr. Phys. Chem.* **29**, 210 (1981).
- [36] U. Maichle, Schülervorstellungen zu Stromstärke und Spannung, *Naturwiss. Unterr. Phys. Chem.* **30**, 383 (1982).
- [37] D. Shipstone, Pupils' understanding of simple electrical circuits. Some implications for instruction, *Phys. Educ.* **23**, 92 (1988).
- [38] J. Closset, Woher stammen bestimmte Fehler von Schülern und Studenten aus dem Bereich der Elektrizitätslehre? Kann man sie beheben?, *Der Physikunterricht*. **18**, 21 (1984).
- [39] T. Wilhelm and M. Hopf, Schülervorstellungen zum elektrischen Stromkreis, in *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendarat und Unterrichtspraxis*, edited by H. Schecker, T. Wilhelm, M. Hopf, and R. Duit (Springer, Berlin, Heidelberg, 2018), pp. 115–138, [10.1007/978-3-662-57270-2_6](https://doi.org/10.1007/978-3-662-57270-2_6).
- [40] P. Tamir, Some issues related to the use of justifications to multiple-choice answers, *J. Biol. Educ.* **23**, 285 (1989).
- [41] F. Haslam and D.F. Treagust, Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument, *J. Biol. Educ.* **21**, 203 (1987).
- [42] A. Chandrasegaran and D. Treagust, The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation, *Chem. Educ. Res. Pract.* **8**, 293 (2007).
- [43] A. Hilton, G. Hilton, S. Dole, and M. Goos, Development and application of a two-tier diagnostic instrument to assess middle-years students' proportional reasoning, *Math. Educ. Res. J.* **25**, 523 (2013).
- [44] P. B. Griffard and J. H. Wandersee, The two-tier instrument on photosynthesis: What does it diagnose?, *Int. J. Sci. Educ.* **23**, 1039 (2001).
- [45] D. R. Sokoloff, Teaching electric circuit concepts using microcomputer-based current/voltage probes, in *Microcomputer-Based Labs: Educational Research and Standards* (Springer, Berlin, Heidelberg, 1996), pp. 129–146, [10.1007/978-3-642-61189-6_7](https://doi.org/10.1007/978-3-642-61189-6_7).
- [46] 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).
- [47] 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).
- [48] S. Zitzelsberger and T. Rabe, Eine Analyse des rhönecktests, in *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam 2010* (LIT Verlag, Münster, 2011), pp. 217–219.
- [49] H. Peşman and A. Eryılmaz, Development of a three-tier test to assess misconceptions about simple electric circuits, *J. Educ. Res.* **103**, 208 (2010).
- [50] S. Hasan, D. Bagayoko, and E. L. Kelley, Misconceptions and the certainty of response index (CRI), *Phys. Educ.* **34**, 294 (1999).
- [51] 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).
- [52] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/mf6b-5xcs> for all the items used in our test instrument.
- [53] K. Schermelleh-Engel and J. C. Gädde, Modellbasierte Methoden der Reliabilitätsschätzung, in *Testtheorie und Fragebogenkonstruktion* (Springer, Berlin, Heidelberg, 2020), pp. 335–368, [10.1007/978-3-662-61532-4_15](https://doi.org/10.1007/978-3-662-61532-4_15).
- [54] D. J. Leiner, Too fast, too straight, too weird: Nonreactive indicators for meaningless data in internet surveys, *Surv. Res. Methods* **13**, 229 (2019).
- [55] A. Kelava and H. Moosbrugger, Deskriptivstatistische Itemanalyse und Testwertbestimmung, in *Testtheorie und Fragebogenkonstruktion* (Springer, Berlin, Heidelberg, 2020), pp. 143–158, [10.1007/978-3-662-61532-4_7](https://doi.org/10.1007/978-3-662-61532-4_7).
- [56] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020103 (2009).
- [57] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, 4th ed., Methodology in the Social Sciences (The Guilford Press, New York, London, 2016).
- [58] B. Groß, Test instrument on simple electric circuits—Raw data and R script for evaluation, 2025, [10.5281/zenodo.16412908](https://doi.org/10.5281/zenodo.16412908).