ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #36845

Statistical Analysis of an Adaptive Concept Inventory in Introductory Electric Circuits for Students and Instructors

Alejandro H Espera

Alejandro is a postdoctoral research associate at the Institute for Advanced Materials and Manufacturing at the University of Tennessee-Knoxville. He has a Ph.D. in Engineering Education and M.A. in Data Analytics and Applied Statistics from Virginia Tech. He also holds B.S. and M.S. in Electronics Engineering. He is an associate professor in the Electronics Engineering Department at Ateneo de Davao University, Philippines. He has done and published research in the areas of additive manufacturing, design of optimized electronic systems, and instructional design and innovations in teaching electrical and electronics engineering core courses.

Nicole P. Pitterson (Assistant Professor)

Statistical Analysis of an Adaptive Concept Inventory in Introductory Electric Circuits for Students and Instructors

ABSTRACT

Concept assessment instruments utilized in electrical engineering education are primarily designed for students by teaching faculty or assessors of the course. Due to the need for identifying inaccurate understanding early at the introductory level, creating a concept inventory that assesses not just students for their knowledge in fundamental electric circuits but also helps novice teaching faculty identify potential inaccuracies in their assumed correct conceptual understanding. This adaptive mechanism is deemed useful for teaching fundamental concepts effectively. Forty-four undergraduate students and teaching faculty from four academic institutions participated in this study. The collected data were analyzed through descriptive and inferential statistical approaches: item-wise difficulty and discriminatory analysis, inter-item correlations, internal consistency reliability using Kuder-Richardson 20 (KR-20) metric, and exploratory factor analysis. Initial findings from the analysis suggested that the instrument attained acceptable validity, yet its reliability could be further improved. The only significant predictor of the scores was the type of participant (faculty or student). In due course, the outcomes of this study will identify which of the test items significantly contribute to achieving the learning of the instrument's electric circuit concept groups and which of the items need to be improved and supplemented. The set of statistical methods selected for this study offers promising means to enhance concept inventories and make them valid and usable for improving the teaching of fundamentals of electric circuits.

Keywords: electric circuits, concept inventory, instrument reliability and validity

1. Rationale

There has been a common understanding that novice instructors (first-career teachers) who teach fundamental engineering concepts tend to explore and try different ways of teaching more than those who have accumulated years of teaching experience to facilitate student learning effectively [1]. Part of this exploration has been to adapt and use existing conceptual assessment instruments. However, the context in which the concepts are situated within the engineering area of expertise is often varied in terms of how abstract or clear they are for students, posing another layer of difficulty for novice instructors [2]. As much as it is crucial to examine inaccuracies in students' understanding of these concepts and the need for them to pinpoint these early on in their study [3], [4], [5], it is also important for instructors that they develop an awareness of their own conceptual knowledge of the content. Thus, a systematic inquiry [1] into this issue is necessary to give instructors, especially beginners, the mechanism to reflect and improve on their teaching strategies [6], [7].

The concept inventory used in this study [8] has been originally designed to comprise two parts: a) 20-item multiple-choice questions involving electric circuits operating in transient D.C. and steady-state A.C., and b) a follow-up structured interview, where the latter is not within the scope of this study. More detail about this instrument is discussed in the methodology section.

The outcome of this analysis will further inform the second part of the instrumentation. This purposeful setup makes this instrument adaptive toward improving the teaching of these introductory concepts because, as a whole, the concept inventory and all its components must work in synergy to be able to achieve our primary goal: the instrument is *to be utilized by instructors who teach introductory concepts in ECE as a tool to self-assess their own conceptual knowledge and that of their students to improve their instruction further*. So, this study specifically aims to achieve the following goals:

Goal 1. *Analyze the difficulty, reliability, and validity of the concept inventory.*

Goal 2. *Identify potential significant factors that affect the test scores.*

Moreover, this study proposes a quantitative methodology, with a chosen set of statistical strategies, into validating assessment instruments such as concept inventories for introductory engineering that can be developed and used across different engineering disciplines.

2. Methodology

This study is based upon our ongoing interest in mitigating inaccuracies in introductory conceptual knowledge among electrical engineering students; thus, we developed our own instrument [8]. There is not much research on the use of concept inventory for instructors because concept inventories, as an assessment tool, are supposed to be designed by the instructors for the students. However, we sought to validate and utilize this instrument adaptively for both students and instructors involved in the learning and teaching of fundamental electrical circuits concepts. This means that the instrument can be used to assess both students and instructors in this specific electric circuit theory, making it adaptive for utility. A quantitative methodology focused on using select statistical tools to provide validation of the instrument was employed to attain the goals of this study.

2.1 Participants

Forty-one (41) undergraduate students and three (3) instructors from electrical and electronics engineering departments across four (4) institutions, 24 from the U.S. and 21 from the Philippines, consented to participate in the study. The general demographics include the participants' institution, year-level, gender, student or faculty, the number of circuits courses taken (students), and the number of years teaching circuits courses (instructors). The medium of instruction used in these institutions is English. See Table 1 for the summary of participants' information.

Table 1. Demographic information

Demographic information	Students	Faculty
Participants	Male (15) Female (27) Not identified (8)	Male (2) Female (0)
Institution	Philippine Institution 1 (21) US Institution 2 (5) US Institution 3 (13) US Institution 4 (6)	
Number of circuits courses taken	1-2 courses: (30) 3-4 courses: (10) 5 or more: (1)	
Number of years teaching circuit courses		2-5 years: (1) 5-10 years: (1) Ten or more: (1)

2.2 Data Collection

The concept inventory was developed [3], [9] to assess knowledge in electrical circuits operating in transient D.C. and steady-state A.C. [8]. These areas of knowledge are determined to be core concepts of electrical engineering across the undergraduate years [10], [11], which means that on a curricular standpoint, the advanced topics are built on these fundamental concepts.

2.2.1 Concept inventory

The concept inventory was administered in the form of a Qualtrics survey to all potential participants throughout the Fall of 2019 and Fall of 2020 semesters. This way, it limits the interaction between the researchers and the participants. The responses of the consenting participants were collected. The concept inventory was proofread at least twice before administering it. Equal weight was assigned for scoring each item. The participants were given at most 40 minutes to respond to the concept inventory.

The 20-item multiple-choice questions have been designed to have zero to minimal numerical content so that no calculator would be needed to respond. Moreover, the specifications in Table 2 show the distribution of questions based on the defined objectives or concept groups.

Table 2. Table of specification (TOS) [8]

Specific objectives (concept groups)	Question item no.
A. Identify and explain a capacitor's behavior.	9, 11, 14, 15
B. Identify and explain an inductor's behavior.	10, 12, 16
C. Explain the process of storing and delivering energy in a first-order circuit with a D.C. source.	17, 20
D. Explain the variables that determine the operation of a first-order circuit's time constant.	4, 6
E. Distinguish between a complete response and forced response in first- and second-order circuits.	2, 3
F. Explain individual behaviors of electric elements in a circuit with an A.C. signal as a source, in steady-state conditions.	1, 7, 8, 14, 19
G. Explain the behavior of an electric device, in terms of power delivery and dissipation, in a circuit with an A.C. source.	5(RMS), 7, 8, 13, 18

3. Motivation and Expected Outcomes

To pinpoint the difficult concepts in the concept inventory, responses were pre-analyzed by looking at the topic-specific items in the 20-item concept inventory, as seen in Table 2. The distribution of items per objective involved a mix of easy and difficult questions, and some items covered more than one objective. The number of participants who managed to get the correct answer for each item was tallied, and then the percentage of these participants according to each objective (A to G in the table) was calculated. It was found that the mean ratio of participants who got the correct answer across objectives was 40%. This number was used to locate the seven objectives wherein percentages fall below 40%. Initial findings suggest that the participants had difficulty attaining objectives B, D, and G. Looking closely, Q5, Q6, and Q8 of the multiple-choice questions (Refer to Appendix) are the most critical questions. In a sense and based on rough estimation and initial findings, the concepts that generally garnered the lowest scores are:

- *The role of time constant in first-order circuit operations.* (Concept Group D)
- *The behavior of reactive elements in terms of power delivery and dissipation in an AC-source circuit.* (Concept Groups B and G)

Interestingly, these significant concepts are also considered difficult based on the findings from a previous study [12] wherein students responded to survey questions about electric circuit concepts and then ranked them based on their perceived importance and level of understanding. The results showed that **transient analysis** and **reactive power** are students' least understood concepts, yet of very high importance to them, which coincided with the circuit concepts identified above. These central concepts are not necessarily independent of the other concepts, such as the behavior of reactive elements in both transient and steady-state conditions. For example, all the other objectives can cover how capacitors and inductors operate. Learning from these initial findings, it is fair to say that teaching the concepts of transients and reactances is challenging [10] because students perceive them as very important, yet, they persistently find these concepts difficult to understand.

By doing statistical analysis, apart from descriptive statistics, we expect that these findings will be further substantiated towards achieving the main goals of this study.

4. Analysis and Results

This section presents the overall analysis of the data, including the summary of the analysis plan, the actual analysis, and the results and discussion of the findings. We used various statistical software throughout the analysis, including SPSS, Minitab, and JMP.

The variables considered for analysis were:

a. *TotalScore* – continuous variable representing concept inventory scores
b. *Type* – categorical variable that tags the participant whether one is faculty or student
c. *Gender* – categorical variable that varies from male, female, to not identified.
d. *Institution* – categorical variable representing four universities
e. *Courses* – categorical variable for the number of years the students have taken circuits courses or teaching years of faculty

4.1 Plan summary

Table 3 presents an overview of the overall analysis plan, including how the techniques used for the following statistical approaches align with the goals of this study.

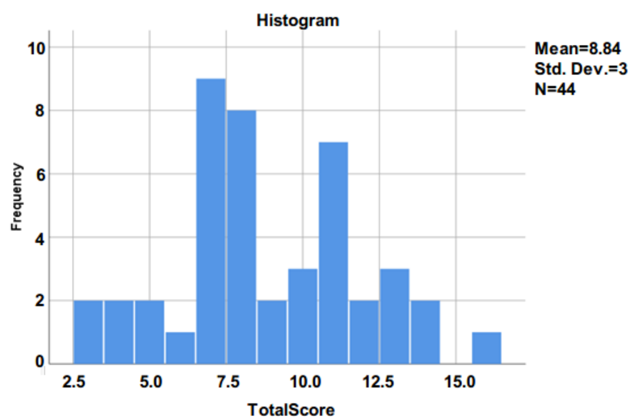
Table 3. Data analysis plan

Goal	Approach	Techniques
1. Analyze the difficulty, reliability, and validity of the instrument.	Descriptive	Central Tendency Measures Normality and Homoscedascity Tests: Shapiro-Wilk's and Normality plots
	Item Analysis	Difficulty Index Discriminatory Index Inter-item Correlations
	Reliability	Internal Consistency Reliability: Kuder-Richardson 20 (KR-20)
	Validity	Confirmatory Factor Analysis
2. Identify potential significant factors that affect the test scores.	Model	Auto Linear Regression: Backward Algorithm

4.2 Descriptives

Descriptives are often used to explore the characteristics of the data set that provide an immediate snapshot or quick observation of the data, informing the steps for further analysis. Here, we characterize the concept inventory scores called *TotalScore* using the following descriptive statistics. On average, the participants' test scores were 8.84 out of 20, with a relatively large variance of 9.3. The lowest and highest scores were 3 and 16, respectively.

Shapiro-Wilk test is one of the many statistical tests used to detect non-normality. The hypothesis assumes that the data is taken from a normally distributed population represented by customarily distributed samples. Here, we see that the p-value is 0.302, which meant that the null hypothesis should not be rejected and assume that *TotalScore* follows normality. On the other hand, given the possibility that the data might be characterized by non-parametric statistic, the Kolmogorov-Smirnov test returned a p-value of 0.011 ($p<0.05$) that led us to reject the null and fail the normality test. However, "significance will be strongly affected by the number of observations, and so only a small discrepancy from normality will be deemed significant for huge sample sizes as may be the case here while huge discrepancies will be required to reject the null hypothesis for small sample sizes" [13]. In other words, large sample sizes are robust to non-normality while small sample sizes may not, according to the Central Limit Theorem [14]. Since we were dealing with a small sample size here, 44 participants, in particular, we could further investigate through normality plots, as shown in Figure 1. The Q-Q plot shows that majority of the data fall along the theoretical normal distribution line. Still, specific data points depart from the line towards the tip, which may agree with the significance found in the Kolmogorov-Smirnov test. However, the plots corroborate with the Shapiro-Wilk results and exhibit overall normality. We investigated normality further in the following sections by fitting several regression models and analyzing the residuals.



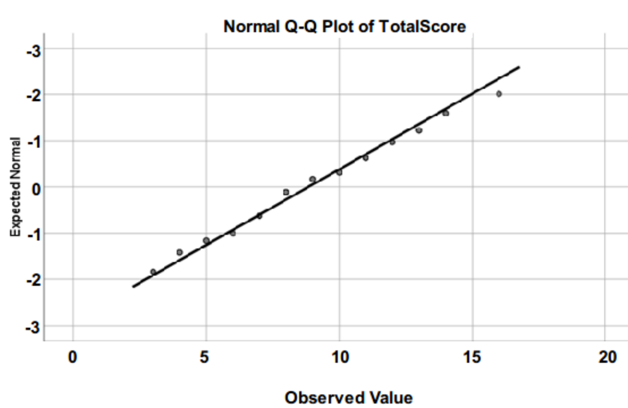


Fig. 1. Normality plots

Normality in the context of this study was crucial as it significantly contributes to the generalizability of the findings from the statistical analysis performed [15].

4.3 Item analysis

Item analysis is a helpful preparatory tool for doing item-wise analysis of an instrument [16] as it provides an overview of how well or poorly the items performed, informing further analysis of factors such as the learning objectives of the test. Two employed metrics include difficulty and discrimination indices of the 20-item multiple-choice questions in the concept inventory. The difficulty index is a standard measure of the difficulty level of a question based upon the respondents' performance. The discrimination index measures how different the top and bottom scorers' performances are for each question [17]. The following formulas [18] are presented here for calculating these indices:

$$Difficulty\ index\ (p) = \frac{No.of\ participants\ with\ correct\ answer}{Total\ number\ of\ participants}\ x\ 100 \qquad \text{Eq. (1)}$$

$$Discrimination\ index\ (D) = \frac{No\ of\ correct\ (top\ 27\%) - No\ of\ correct\ (bottom\ 27\%)}{No\ of\ participants\ (27\%\ of\ total)} \qquad \text{Eq. (2)}$$

Decisions were based on specific acceptable ranges provided in Table 4. This matrix was used to check each item's difficulty and discrimination performances [16]. Then, in Table 5, the items that needed further investigation were labeled accordingly in red font (Items 1, 2, 3, 5, 6, and 8). Those highlighted in blue have acceptable difficulty levels and have discriminated well based on p>0.20 and discrimination D>.25.

Table 4. Decision matrix for difficulty and discrimination indices [16]

Metric	Range	Interpretation	Action
Difficulty (p)	0 to 0.25	Difficult	Revise/discard
	0.26 to 0.75	Right difficulty	Retain
	0.76 above	Easy	Revise/discard
Discrimination (D)	-1.0 to -.05	Can discriminate but the item is questionable	Discard
	-0.55 to 0.45	Non-discriminating	Revise
	0.46 to 1.0	Discriminating item	Retain/include

Table 5. Difficulty and discrimination indices per item

Item	Total correct	No. Correct (top 27%)	No. Correct (lower 27%)	*Difficulty (p)*	*Discrimination (D)*
1	15	5	4	0.34	0.08
2	18	7	5	0.41	0.17
3	13	5	3	0.30	0.17
4	27	12	4	0.61	0.67
5	7	2	1	0.16	0.08
6	9	3	1	0.20	0.17
7	18	7	3	0.41	0.33
8	9	3	1	0.20	0.17
9	17	8	2	0.39	0.50
10	17	7	2	0.39	0.42
11	31	12	4	0.70	0.67
12	30	11	4	0.68	0.58
13	21	9	4	0.48	0.42
14	17	9	3	0.39	0.50
15	22	9	5	0.50	0.33
16	13	6	3	0.30	0.25
17	27	10	3	0.61	0.58
18	20	8	3	0.45	0.42
19	39	12	7	0.89	0.42
20	19	6	3	0.43	0.25

4.4 Reliability

One of the hallmarks of a good assessment instrument is its reliability [19], [20]. A reliable instrument means it can measure a single cognitive factor. In the case of this concept inventory, its internal consistency reliability tells how well the test, across the items, measures the conceptual knowledge on steady-state A.C. and transient D.C. Although this concept inventory involves multiple concept groups, as stated in the learning goals, they all converge to the umbrella knowledge base. With a reliable instrument, consistency and reproducibility are observed.

Measures of reliability differ in terms of the type of instrument or questions used and the type of collected data. Cronbach's alpha is commonly used for survey questions of Likert's scale type. However, for dichotomous data, such as Yes or No, or questions with one correct answer from among the choices, Kuder-Richardson 20 or KR-20 index is an appropriate measure of internal consistency reliability [21], [22], which is fundamentally a version of Cronbach's alpha for a set of dichotomous question. So, in this part of the analysis, KR-20 was used because the concept inventory involves multiple-choice questions with one correct answer.

The summary statistics showed a variance of .034, which was not as spread out from the mean. The reliability test revealed a KR-20 of 0.558, which was a little below good test reliability of 0.60. Certain factors might have contributed to this result; for instance, the higher the number of test items, the more heterogeneous the participants were, and the less difficult the questions were, leading to higher test reliability. In the further section of this paper, improving this index will be explored by considering the factors mentioned. We also explored taking the test scores and grouping them based on the learning objectives or concept groups and then testing the reliability again. Overall, we attained a Cronbach's alpha of 0.541, which was not significantly different from the value we obtained from the reliability when accounting for each item.

4.5 Validity

A valid instrument means that it has the acceptable ability to measure what it is supposed to measure. When using existing instruments in social science, a validated instrument means it is ready to be utilized for its purpose [22]–[24]. To explore the validity of this concept inventory, exploratory and confirmatory factor analyses could be used. Exploratory factor analysis was more appropriate because, although the concept groups or learning goals have already been set and the potential predictors to test scores have already been identified from the data set, we could explore whether the latent concepts align with the concept groups in our defined learning objectives. Factor analysis uses algorithms and statistical tools to evaluate the evidence whether the analytical findings and the hypotheses make sense. The purpose of performing factor analysis for this concept inventory was to see whether the concept groups that were initially defined were seen as underlying factors to the test scores from all the items in the test. Also, we were interested to see whether the predictors such as institution, gender, type of participant, and courses taught or taken, were seen as significant factors in the test scores. However, this notion may be farfetched as they are straightforward variables. Hence the name exploratory, this factor analysis explored these presumptions, or shall we state it in a way that these factors might be considered latent variables (or indirectly measurable constructs such as particular ECE conceptual units) for which this analysis could explore their existence based on evidence, and then provides validity into what was categorically measured by the instrument.

The exploratory factor analysis used in this section was principal component analysis as the main algorithm and Varimax with Kaiser Normalization for the rotation method. So, the term “factor” has been used interchangeably with “components” in this discussion. With scree plot analysis, eight different latent constructs were seen from the data set, which needed to be verified. The rotated component matrix showed us that the factors were not entirely independent, and some components had negative correlations with the items. However, with careful cross-examination with the predefined learning objectives or concept groups in Table 2, this analysis showed expected conceptual relationships within and across concept groups. These concepts are *inductor behavior, time constant, capacitor behavior, energy storage, power, circuit components, first-order circuits, and second-order circuits*. However, because these introductory ECE concepts posed dependence with each other and were part of the overarching steady-state A.C. and transient D.C. umbrella construct, these concepts indicated as components had confirmed the predefined concept groups.

To further establish the relationship of the concept groups with the overarching conceptual knowledge of steady-state A.C. and transient D.C., another round of factor analysis was performed for the seven objectives or concept groups. The results showed two major components. The analysis confirmed that the objectives constitute questions that deal with (1) AC/DC circuit analysis as a whole and (2) an independent analysis of the electrical device, hence the two components.

4.6 Modeling

It is also of great interest that predictors of the test scores were explored for significance. A linear regression model was created and fitted against the data set using a backward algorithm to predict test scores based on the type of participant, institution, gender, and number of courses. Although these predictors were categorical, they were of nominal form and were coded accordingly.

Table 6 presents a correlation table across the variables of interest. The table shows a weak correlation between the response and predictor pairs. However, significant relationships were seen on the following (highlighted in blue):

- Total score and type of participant (faculty or student) with $p<0.001$
- Type of participant and number of courses taken or taught with $p<0.05$
- Gender and institution with $p<0.001$

Overall, it could be anticipated that these findings may create interaction effects between predictor variables. However, the goal of the model was to predict which of the independent variables affected the total scores significantly. Based on these findings, the type of participant, whether students or faculty, had a statistical significance with the total scores. Further investigation were performed to prove that this was, in fact, evident in the data set.

Table 6. Correlations across variables

		Total score in the concept test out of 20	Type of Participant	Name of institution	Gender of Participant	Number of courses taken (students) or taught (faculty)
Pearson Correlation	Total score in the concept test out of 20	1.000	.402	-.076	.019	.193
	Type of Participant	.402	1.000	-.126	-.197	.317
	Name of institution	-.076	-.126	1.000	.362	.090
	Gender of Participant	.019	-.197	.362	1.000	-.131
	Number of courses taken (students) or taught (faculty)	.193	.317	.090	-.131	1.000
Sig. (1-tailed)	Total score in the concept test out of 20	.	**.003**	.313	.451	.105
	Type of Participant	**.003**	.	.208	.100	**.018**
	Name of institution	.313	.208	.	**.008**	.280
	Gender of Participant	.451	.100	**.008**	.	.199
	Number of courses taken (students) or taught (faculty)	.105	**.018**	.280	.199	.

After performing a backward linear regression, the fourth model best described the data set in terms of goodness of fit and most significant predictors. As shown in Table 7, Model 1 considered all the predictors into the model, and then one by one, a predictor was removed based on the results of the defined criterion until Model 4 was achieved. Model 4 showed that the type of the participant was indeed the only significant predictor of the test scores, which corroborated with the previous findings from the correlation analysis. The very small R-square of 0.162, or 0.142 adjusted (see Table 7), said much about the current data set. This meant that the model only described 16.2% of the variance in the data. The main reason for this is that a larger sample size may be needed, especially since there were only three faculty and 41 student participants – and the type of participant is found to be a statistically significant factor in the test scores.

Table 7. Model summary[e]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.429[a]	.184	.100	2.900
2	.421[b]	.178	.116	2.875
3	.414[c]	.172	.131	2.849
4	**.402[d]**	**.162**	**.142**	**2.832**

a. Predictors: (Constant), Number of courses taken (students) or taught (faculty), Name of institution, Type of Participant, Gender of Participant
b. Predictors: (Constant), Number of courses taken (students) or taught (faculty), Type of Participant, Gender of Participant
c. Predictors: (Constant), Type of Participant, Gender of Participant
d. Predictors: (Constant), Type of Participant
e. Dependent Variable: Total score in the concept test out of 20

The ANOVA results of the four models created from the auto linear regression using a backward algorithm showed that although Models 2 and 3 were significant at $p<0.05$, Model 4 was significant at $p<0.001$, and it has an $F\text{-}statistic>4$, which is twice the Model 3's. Thus, Model 4 performed best among the four models for the purpose we defined.

Analyzing the residuals to investigate normality further, it was observed that its standardized residuals generally follow normality assumptions with the total test score as a response. This means that regression modeling was an appropriate modeling technique.

4.7 Improving test reliability

Based on difficulty level and discrimination index results from section 4.3, items with questionable difficulty and discrimination (Questions 1, 2, 3, 5, 6, and 8) according to criteria difficulty $p>0.20$ and discrimination $D>.25$ were removed from the test set of questions. Another round of KR-20 analysis was performed, and the internal consistency reliability has indeed increased from 0.558 to 0.610, which is now above the acceptably good index of 0.60.

Several relevant inferences were made from the findings of all the analyses performed in this study. The following section provides a summary of the discussion and the conclusion.

5. Discussion and Conclusion

This study's primary goals were to analyze the difficulty, reliability, and validity of the 20-item concept inventory for introductory electric circuits and identify potential significant factors that affect the test scores among the faculty and student participants. And these goals were met accordingly. The item analysis findings showed that some challenging and non-discriminatory items must be further investigated to ensure that these items were not immediately accounted for as a misconception or inaccurate conceptual knowledge. Reliability tests showed that internal consistency could be improved by temporarily removing the questionable items identified in the item analysis. However, this finding did not necessarily mean that the items were unreliable or invalid. It only meant that these items that fall under specific objectives of the test needed to be supplemented with other support mechanisms or other forms of assessments. The exploratory factor analysis results showed how specific concepts under the umbrella of steady-state A.C. and transient D.C. were mostly interdependent and not necessarily mutually exclusive constructs. This was clearly demonstrated by the principal component analysis for both item-wise and objective-wise concepts. The auto linear regression using a backward algorithm produced the best possible model based on the defined constraints that the type of participant (faculty or student) was the only significant predictor in the concept inventory scores. This meant that the overall test scores of students versus faculty were sufficiently distinguishable given the data set. However, the very low goodness-of-fit measures suggested that more samples might be needed to make the inference more substantial. Most significantly, the data needed more representation among faculty. However, more student participants might also be needed to conclude that the findings are more generalizable to a broader population.

The "adaptive" nature of this instrument proposes to have two parts, the actual concept test and the semi-structured interview. The latter was not a part of the scope of this study. However, validating the concept test will further enhance the second part of the instrument, for which the overall goal is to allow a deep probe of their responses to the concept questions. Researchers will find this to be an interesting addition to a purely quantitative assessment instrument as it contextualizes the responses with participants' knowledge constructions and mental models. Moving forward, the developed instrument can be helpful for education researchers and educators, especially novice instructors, who have a particular focus on assessing complex introductory concepts, which can be drafted in the form of similarly structured concept inventory such as the one used in this study but intended for different engineering, or even science, disciplines with the use of the selected and proposed statistical approaches and quality measures employed in this study. More importantly, through the methods of assessment of concepts presented in this study, novice instructors can have opportunities to adopt them for their own reflective practices towards improving their approaches to teaching these concepts.

References

[1] National Research Council, "How experts differ from novices," in *How People Learn: Brain, Mind, Experience, and School*, Washington, D.C.: National Academies Press, 2000, pp. 31–50. doi: 10.17226/9853.

[2] D. A. Muller, J. Bewes, M. D. Sharma, and P. Reimann, "Saying the wrong thing: improving learning with multimedia by including misconceptions," 2007, doi: 10.1111/j.1365-2729.2007.00248.x.

[3] D. L. Evans *et al.*, "Tools for assessing conceptual understanding in the engineering sciences," *Frontiers in Education Conference*, p. 7803, 2002, doi: 10.1109/FIE.2002.1158151.

[4] T. Pinarbaşi, N. Canpolat, S. Bayrakçeken, and Ö. Geban, "An investigation of effectiveness of conceptual change text-oriented instruction on students' understanding of solution concepts," *Research in Science Education*, vol. 36, no. 4, pp. 313–335, 2006, doi: 10.1007/s11165-005-9003-4.

[5] S. Bull, T. J. Jackson, and M. J. Lancaster, "Students' interest in their misconceptions in fi rst-year electrical circuits and mathematics courses," *International Journal of Electrical Engineering Education*, vol. 47, no. 3, pp. 307–318, 2010, doi: 10.7227/IJEEE.47.3.6.

[6] A. H. Espera and N. Pitterson, "Teaching circuit concepts using evidence-based instructional approaches: A systematic review," 2019.

[7] N. Pitterson and R. Streveler, "Teaching and learning complex circuit concepts: An investigation of the intersection of prior knowledge, learning activities, and design of learning environments," 2016. doi: 10.18260/p.26018.

[8] A. H. Espera, N. Pitterson, and R. A. Soto Perez, "Understanding potential misconceptions shared between instructors and students in fundamental electric circuits," 2020.

[9] R. A. Streveler, R. L. Miller, A. I. Santiago-roma, M. A. Nelson, M. R. Geist, and B. M. Olds, "Rigorous Methodology for Concept Inventory Development : Using the 'Assessment Triangle' to Develop and Test the Thermal and Transport Science Concept Inventory (TTCI)*," *International Journal of Engineering Education*, vol. 27, no. 5, pp. 968–984, 2011.

[10] J. Bernhard and A. Carstensen, "Learning and teaching electrical circuit theory," *Physics Teaching in Engineering Education*, 2002.

[11] A. K. Carstensen and J. Bernhard, “Student learning in an electric circuit theory course: Critical aspects and task design,” *European Journal of Engineering Education*, vol. 34, no. 4, pp. 393–408, 2009, doi: 10.1080/03043790902990315.

[12] R. A. Streveler and K. A. Smith, “Conducting Rigorous Research in Engineering Education,” *Journal of Engineering Education*, vol. 95, no. 2, pp. 103–105, Apr. 2006, doi: 10.1002/j.2168-9830.2006.tb00882.x.

[13] The British Academy Centre for Multilevel Modelling, *Checking for Normality*. 2020.

[14] G. Byrne, “A Statistical Primer: Understanding Descriptive and Inferential Statistics,” *Evidence Based Library and Information Practice*, vol. 5, no. 4, pp. 21–38, 2006, doi: 10.1086/667435.

[15] J. W. Cresswell, *Research Design: Qualitative, Quantiative, and Mixed Methods*. Thousand Oaks, California: SAGE Publications, Inc., 2014.

[16] “Test and Item Analysis « Chi Tester Users’ Manual.” https://chitester.wordpress.com/instructor-guide/section-4-reporting-features/test-and-item-analysis/ (accessed Nov. 18, 2021).

[17] “Classroom Assessment | Basic Concepts.” https://fcit.usf.edu/assessment/selected/responsec.html (accessed Nov. 18, 2021).

[18] “Calculating Difficulty, Discrimination and Reliability Index/Standard Error of Measurement | PPUKM.” https://ppukmdotorg.wordpress.com/2015/04/02/calculating-omr-indexes/ (accessed Nov. 19, 2021).

[19] N. Golfashani, “Understanding reliability and validity in qualitative research,” *The Qualitative Report*, vol. 8, no. 4, pp. 597–607, 2003.

[20] N. E. Canney and A. R. Bielefeldt, “Validity and Reliability Evidence of the Engineering Professional Responsibility Assessment Tool,” *Journal of Engineering Education*, vol. 105, no. 3, pp. 452–477, 2016, doi: 10.1002/jee.20124.

[21] D. Reliability, “Handout 4 : Establishing the Reliability of a Survey Instrument,” pp. 1–23, 2016.

[22] J. W. Cresswell, “Quantitative methods,” in *Research design: Qualitative, quantitative, and mixed methods approaches*, Thousand Oaks, CA: Sage Publishing, 2009, pp. 145–169.

[23] C. K. Y. Chan, Y. Zhao, and L. Y. Y. Luk, “A Validated and Reliable Instrument Investigating Engineering Students’ Perceptions of Competency in Generic Skills,” *Journal of Engineering Education*, vol. 106, no. 2, pp. 299–325, 2017, doi: 10.1002/jee.20165.

[24] J. Zhu, Y. Li, M. F. Cox, J. London, J. Hahn, and B. Ahn, “Validation of a survey for graduate teaching assistants: Translating theory to practice,” *Journal of Engineering Education*, vol. 102, no. 3, pp. 426–443, 2013, doi: 10.1002/jee.20014.