RELIABILITY OF A CONCEPT INVENTORY TO DETERMINE THE LEVEL OF STUDENTS IN STATICS

Carlos Andrés Restrepo García
Deparment of Mechanical Engineering
Eafit University
Medellín, Colombia
crestr19@eafit.edu.co

Jaime Leonardo Barbosa Pérez
Deparment of Mechanical Engineering
Eafit University
Medellín, Colombia
jbarbosa@eafit.edu.co

Jorge Luis Restrepo Ochoa
Deparment of Mechanical Engineering
Eafit University
Medellín, Colombia
jrestrep@eafit.edu.co

***Abstract*—Basic core courses in engineering are often taught and assessed only through procedural approaches, neglecting the conceptual grounds of the subject. In an effort to change that, a Concept Inventory test was applied to 195 students at Eafit University enrolled in Statics. All of the students took the test inside in the campus facilities and were monitored during the whole session to make sure they did not share information. The students took a Computer-Based Test of the Concept Inventory. Three new items were added to the original test of 27 questions, in order to improve the reliability of two specific groups of concepts. For the test was found a mean of 9.7 and a standard deviation of 5.28. For the overall test, the Cronbach's alpha coefficient was 0.81. The psychometric analyses show that all of the items in the test, except one, present an appropriate fit for discrimination and difficulty parameters of the test. Item 26 was deleted when performing the analysis of sub-scales reliability because presented abnormal value for difficulty. For the cluster of Equilibrium, it was seen that deleting this item resulted in an increase of the alpha from 0.2 to 0.36. For the sub-scale of Free-Body Diagram was seen a decrease in the reliability when two new items were added to it. All of the other sub-scales presented good reliability, most of them, above 0.5. The item-person map shows that the overall latent trait of students is lower than the overall test difficulty, meaning that most of the students found the test difficult. One of the conclusion to be drawn from the results is that the students subjected mainly to procedural approaches in teaching, do not perform well in conceptual tests. The results support the claim that Statics can be seen as a collection of concepts that can be clustered in independent groups for teaching.**

***Index Terms*—concept inventory, Statics, engineering education.**

I. Introduction

The teaching strategies in engineering courses are mainly focused on procedural approaches and not conceptual. Although conceptual understanding is vastly studied in sciences, there is a lack of research focusing on the engineering sciences [1].

In the teaching of mathematics there exists results showing that students who received conceptual-based classes performed better in mixed conceptual and procedural assessments compared to those students who only received procedural-based instruction [2]. In the engineering education, one of the efforts made to address the lack of conceptual understanding are the well-known Concepts Inventories (CI's). The CI's are a kind of test instrument specially designed to assess the conceptual understanding of students in a specific area of knowledge, as well as to identify the most common errors made by students when being asked about concepts. The Statics Concept inventory of professor Steif [3] consists in a set of 27 questions that aims to correctly measure the conceptual understanding in statics, and it has been applied with satisfactory results in both students and professional engineers in companies [4], [5].

In the literature can be found examples in the applications of CI's in the engineering courses, like the ones cited above and the famous Force Concept Inventory [6], nevertheless, no applications of this kind of tests are found in Spanish. This paper aims to show and to compare the results in the application of a CI's in a university in Colombia, where the teaching is merely procedural, to those existent results of the CI's in the US.

II. Methodology

A. The Test

The original test from professor Steif was translated into Spanish. All of the 27 questions of test were used. The test is clustered into four groups of concepts and it is divided in nine groups as it follows: forces on collection of bodies, Newton's third law, static equivalence, roller, slot, negligible friction, representation, friction and equilibrium. In a previous analysis was determined that item 26 had an unacceptable difficulty and discrimination value [7], also that items 11, 21 and 26 had alpha-with-item-deleted. in order to improve the reliability of the test,3 new items were added to the original test completing a 30 questions test. The 3 new questions were formulated keeping the structure of the original items and their incorrect responses were also thought in relation to the most common errors made by students. All of the items in the test

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

were digitalized so a Computer-based Test could be done. The computer-based version of test was set to last 60 minutes and it was updated to a virtual platform implemented in Moodle, all of the students had access to the platform from day one of classes. The test could be taken during whole week previous to the end of the course but it was only visible for the students at the place assigned for this. Students could only take the test once and they were supervised during the exam.

B. New Items in the test

- *Item 28:* The bar is subjected to two forces with magnitudes known and the forces are acting in the directions shown.

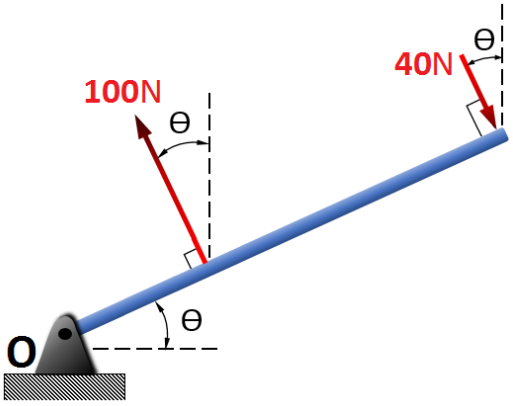

Fig. 1. Additional item in the test – Item 28

For the previous system, what is the free body diagram that best represent the forces exerted by the pin in O?

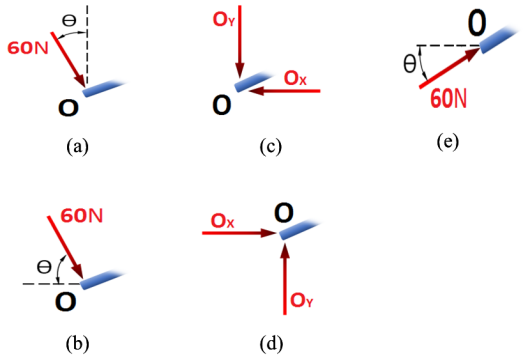

Fig. 2. Response options for Item 28

- *Item 29:* The system shown consist in a disc subjected to a moment of magnitude M and a break that acts due to force P.

  For the previous system, what is the best free body diagram that best represents the forces of the disc on the break?

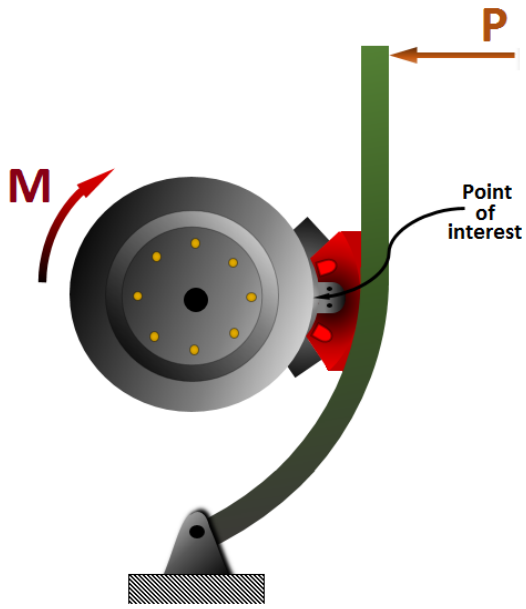

Fig. 3. Additional item in the test – Item 29

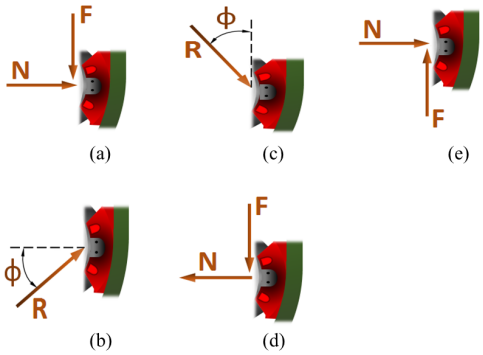

Fig. 4. Response options for Item 29

- *Item 30:* In the system shown, the weight of the block compresses the spring from its initial position until a value "x" without reaching its maximum compression

  Having in mind the previous conditions, what is the free body diagram that best represents the forces acting on the block?

C. Participants

Two weeks before the end of the semester an email was sent to all students enrolled in Statics inviting them to take the conceptual test. The assistance to the test was voluntary but those students who took the test obtained a bonus in their grades for the final examination. The bonus obtained was proportional to the results in the CI. The suggested

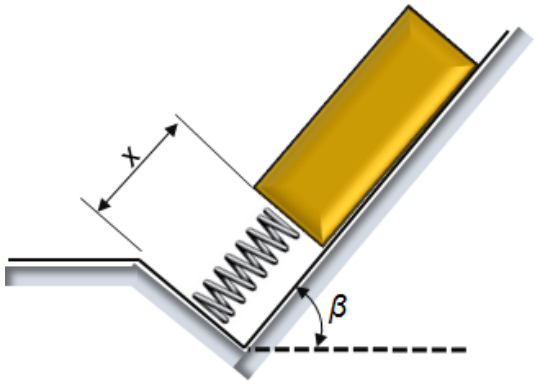

Fig. 5. Additional item in the test – Item 30

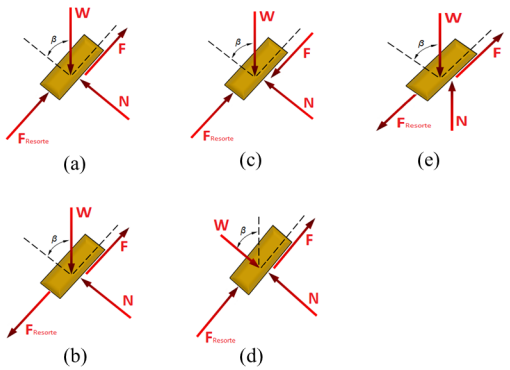

Fig. 6. Response options for Item 30

schedule in which they could take the test was attached to the email. The test was taken by a total of 195 students at Eafit university facilities and it was monitored by the coordinator of Statics course along with a master student to make sure that the students did not share information during the test. From all of the students, females were the 27% of the sample (52 students) and males were the 73% (143 students). There were no considerable ethnic distinctions in the group. The participants were all students of either mechanical, production or civil engineering.

III. Results

A. Descriptive statistics

In order to have a better comparison with previous results of the CI, the test was first analyzed counting only the 27 original questions. From this first analysis, it was obtained a mean and standard deviation of 8.5, and 4.9 respectively, the maximum score was 25 and the minimum was 0, the median was 8. The mean and SD were for males 9.1 and 5.1, and for women 6.7 and 3.9, respectively. For the 30 items test was found that the overall test's mean and SD was 9.7 and 5.28 respectively. The Cronbach's alpha coefficient was 0.81 which shows good evidence of reliability of the test. It is also very close from the results obtain by the author [4]. They are also very similar to the results obtained in a framework used to evaluate concept inventories [7].

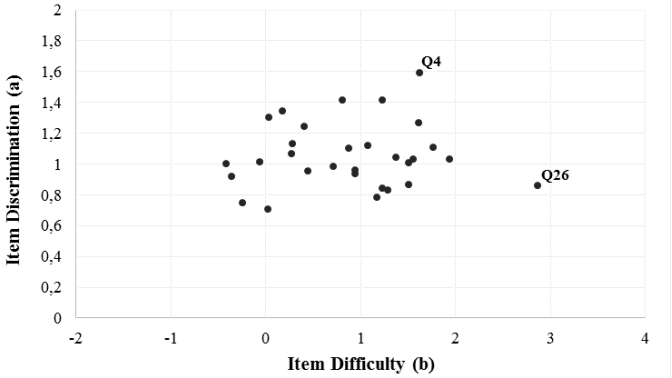

Fig. 7. Item Discrimination vs Item Difficulty

B. Psychometric Analysis

The psychometric analysis was performed using Xcalibre 4.2. The estimation method used for the analysis was the Maximum Likelihood Estimation (MLE). The test was considered as a dichotomous test. The 2PL model was used. The parameter "a" refers to the discrimination power that each item possesses. A high discrimination value means that, a specific item is better at distinguish or separate students with high values of the latent trait (θ) to those with low values. As it can be seen in Fig.7, for the test, it was only found positive discrimination values but it was possible for it to have negative ones.

All of the items in the test fell within the common accepted range of discrimination, which is, only positive values above 0.2. The values of the b parameter refer to the difficulty of each item and they were measured in a -4 to 4 scale, items with more positive b parameters are more difficult for test takers. It is the location in the theta continuum where the probability of endorsing an item equals 0.5. The test comprised items with easy, medium and hard levels. The item 26 is found to have a higher b value than the rest of the items, it suggests that the estimation of the IRT parameters did not have a good fit for this item. The abnormal behavior of this item was reported previously [7], Q26 had a and b values out of range. This item should be reviewed or not included in future tests.

In the Fig.8 can be seen that the overall difficulty of the test is pushed to the right of the overall ability or latent trait of students taking the test. The students' ability and the difficulty are measured in the same scale. For this specific case, the graph indicates that in general terms the test was considered hard by the students.

Sub-scales reliabilities are presented in table I, it can be seen that there is a significant increase in the reliability value for the equilibrium cluster when removing the problematic item and including the new one, however, this value of .36 is still low

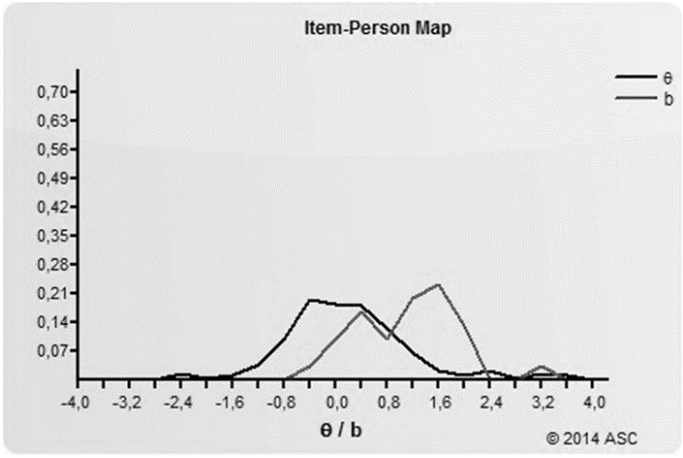

Fig. 8. Item-Person map

for a low-stakes reporting purposes. It can be seen that adding 2 items to the free-body diagram category resulted in a poorer correlation of the category, it can be due to the fact that the new two items were not that specific in the concept of drawing the forces between a collection of bodies but they both needed instead the use of wider concepts like friction. In general, the other categories present reasonable alpha's (above 0.5) indicating good reliability for low-stakes reporting purpose.

C. Tetrachoric correlation

The tetrachoric correlation is a measure of association between two variables that have been measured in a dichotomous way, even thought, those variables are not necessarily binary in real life [8]. The tetrachoric coefficients for each of the questions in the CI was obtained in R using the library cited in [9]. The matrix was plotted in a heat-map shown in Fig.9.

The darkest points in the figure indicate that there is not a strong correlation between those items, the white points show a strong correlation between those items. It can be seen that

TABLE I
CRONBACH'S ALPHA FOR ORIGINAL AND MODIFIED TEST

Original test			**Modified test**	
Category	***n***	α	***n***	α
Free-body diagram	3	0.62	5	0.49
Newton's third law	3	0.67	3	0.67
Static equivalence	3	0.49	3	0.49
Roller	3	0.67	3	0.67
Slot	3	0.74	3	0.74
Frictionless point contact	3	0.54	3	0.54
Representation	3	0.38	3	0.38
Friction	3	0.59	3	0.59
Equilibrium	3	0.20	3	0.36

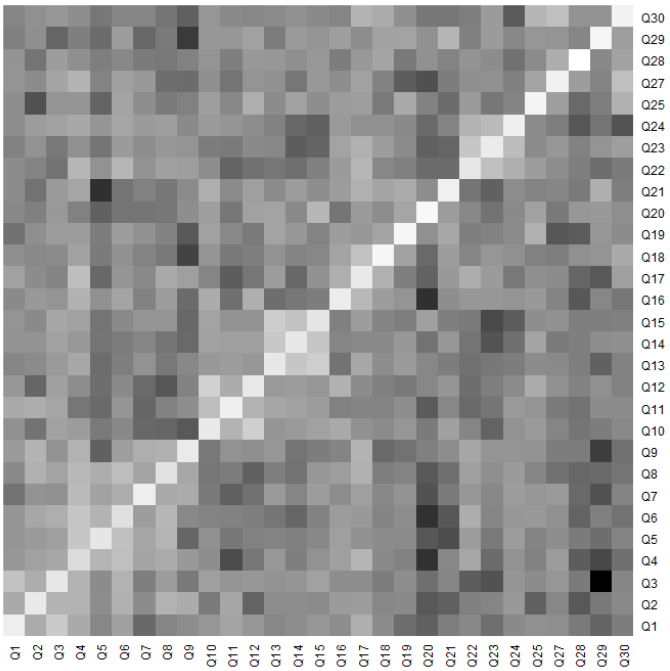

Fig. 9. Heat-map of tetrachoric correlation

there is a clear tendency of clustering in groups of three items in the diagonal, which suggests that the items correlate with the other items within the initial categories suggested by the author.

IV. Discussion and conclusions

It is important to design better strategies for teaching so the professor can be sure that the knowledge he is willing to transmit is in fact being held on students' minds. It can be concluded from this paper that students from the Eafit University find it difficult to solve conceptual problems in Statics. One possible explanation for this is that teaching in the University is almost completely procedural oriented. The examinations and quizzes are merely practical problems where the student needs to solve it making use of procedures to obtain the correct answer. From the Concept Inventory can be said that it presents good reliability and that most of the items possess reasonable parameters of both difficulty and discrimination which reflects a correct design of the test and it supports the claim that statics can be divided into sub-groups of concepts that although keep relation with one another can be taught and evaluated separately. The category of equilibrium is found to be problematic, It can be due to the different ways of structuring these last three questions, although all of them ask about equilibrium, the way to ask is very different in all of three, which is not the usual structure throughout the test.

Acknowledgments

Paul Steif thanks for making available the Concept Inventory and for the early responses in the emails.

REFERENCES

[1] R. A. Streveler, T. A. Litzinger, R. L. Miller, and P. S. Steif, "In the engineering sciences: Overview and future research directions," *Journal of Engineering Education*, no. July, pp. 279–294, 2008.

[2] J. Engelbrecht, A. Harding, and M. Potgieter, "Undergraduate students performance and confidence in procedural and conceptual mathematics," *International Journal of Mathematical Education in Science and Technology*, vol. 36, no. 7, pp. 701–712, 2005.

[3] P. S. Steif, "Initial data from a statics concept inventory," in *Proceedings of the 2004 American Society of Engineering Education Conference & Exposition*, 2004. [Online]. Available: http://link.aip.org/link/abstract/ASMECP/v2004/i47233/p35/s1%5C nhttp://www.andrew.cmu.edu/user/steif/papers/ASEE_Statics_CI_Paper.pdf

[4] P. S. Steif and J. A. Dantzler, "A statics concept inventory: development and psychometric analysis," *Journal of Engineering Education*, no. 4, pp. 363–371, 2005.

[5] O. Ha and S. Brown, "A factor analysis of Statics Concept Inventory data from practicing civil engineers," *Proceedings - Frontiers in Education Conference, FIE*, vol. 2016-Novem, pp. 1–4, 2016.

[6] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *The Physics Teacher*, vol. 30, no. 3, pp. 141–158, 1992. [Online]. Available: http://aapt.scitation.org/doi/10.1119/1.2343497

[7] 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," *Journal of Engineering Education*, vol. 104, no. 4, pp. 454–496, 2015.

[8] K. Pearson, "Mathematical contribution to the theory of evolution. VII. On the correlation of characters not quantitatively measurable," *Philosophical Transactions of the Royal Society*, vol. 195, pp. 79–150, 1990.

[9] W. Revelle, "psych," *Procedures for Personality and Psychological Research*, 2018. [Online]. Available: https://cran.r-project.org/package=psych Version = 1.8.3.