Assessment of the Effects of Compulsory Online Learning During Pandemic Time on Conceptual Knowledge Physics

[1] Elmira Kushta, [2]Dode Prenga, [3]Silvana Miço, [4]Polikron Dhoqina,

1 Department of Mathematics, Faculty of Technical Sciences, University "I.Qemali" of Vlora 1elmira.kushta@univlora.edu.al

2dode.prenga@fshn.edu.al, 3Silvana.mico@fshn.edu.al, 4 polikron.dhoqina@fshn.edu.al,

Article Info
Page Number: 6382-6391
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022

Abstract

Herein we have considered a semi-analytic insight on effects of the online learning limitation during pandemic closure on the high school student's knowledge in physics. The acknowledgment of the effects is realized by conducting standard concept inventory tests and analyzing their raw results and accompanying specifics features. The Rasch technique is used for assessment of question's difficulties and student's abilities to solve standard tests by which we identified the learning problems and drawbacks. After establishing frequent errors from the concept inventory test answers and evidencing the indecisiveness in responding questions, we used the Likert 5-scale instrument for a more detailed analysis. We concluded that the lack of demonstration and live laboratory work supporting physics teaching is the most influential factor for the temporal drawback observed.

Keywords: Concept inventory, physics education, Rasch analysis, Likert scale.

I. INTRODUCTION

The full-scale online education imposed during pandemic closure has caught the education system by surprise in many countries. Those extraordinary circumstances are expected to have affected the efficiency of science knowledge transmission through education system, which motivated us to consider a direct measurement of conceptual knowledge in physics on the students that have got lectures by the full online mode. Nevertheless, we have observed from the public education system sources of data that no significant change of the exams outcomes has been recorded for the 2020-2022, therefore to investigate possible problems and shortcomings we preferred to measure directly the knowledge level by employ calibrated instruments. Technically, the evaluation of the specific knowledge features can be achieved through procedural and conceptual tests [1]. So, the conceptual knowledge is focused on the understanding of concepts and relationships between variables, or definitions and principles, whereas the procedural knowledge considers the ability to solve step by step problems, the fluency in the application of the adequate algorithm in the judgment, see [2], [3], [4], etc. However, in many situation, procedural or routine exams might hide a thorough knowledge inventory measurement. In this case, the dictated conceptual knowledge investigation is expected to be particularly more efficient. The measurement of the conceptual knowledge in practice can be realized by the Concept Inventory assessment [5], [6], [7], and accordingly, standardized CI tests for those evaluations are available and freely accessible. In those typical analysis, the accuracy of measurements and consistency of the interpretations must be guaranteed, therefore we use Rasch calibration techniques throughout the calculations and data elaboration. The Rasch analysis is a well-known psychometric method to ascertain calibrated sociometric assessment, but moreover, it is utilized also to evaluate key important parameters such as the difficulty of a test' item, the student ability to solve the test, guessing behavior in checking correct answer etc. Note that CI test are basically multiple-choice enquiries, and potentially a student might chose correct answer without solving it. The calculation of the estimated probability for the student to solve a certain question powered by the Rasch procedure helps to identify those hidden behavior. We can get more information by using n-scales instruments as provided by Likert techniques, and we used it also in this work.

II. DATA AND METHODS

The raw data are gathered from the CI test undertaken on groups of students that are considered as "sample" in statistical point of view. The solutions provided by those data are converted initially onto binary variable by assigning 1 or 0 for correct or incorrect answer respectively. Those new data are stored in a table R_{mxn} of the students (m) records as rows and (n) tests items as columns. It is considered as the raw data of probabilities that student (i) can solve item (j). Next, based on the Rasch procedure that we will introduce below, we calculate the quantities for analyzing further, as item difficulties, expected probabilities to solve the test, outliers, adjusted CI scores etc. The calculation of the item's difficulty and students' ability is made by the Rasch technique described in [8], [9], [10], etc. The tests purposed for general assessment herein belongs to the standard CI tests as the Force Concept Inventory introduced in [2], the motion CI (MCI) test, and the electromagnetic CI test (EMCI) discussed in [1]. Also, we conducted dedicated testing for specific analysis, as the evidence of the contextual issues related to the laboratory works use to support lecturing in physics, for identification of the dominant factor causing a physics test failure and for analysis of the ambiguity in giving a specific answer. So, we have realized a simplified FCI test (SFCI) by reducing the number of alternatives and employing easier contextual conditions on the problems statement. Next, for casualty analysis we have conducted the FCI-like test composed by a physics and a mathematics section. Herein, the physics part was based on ideas of the references [11], [12], [7], and on our teaching experience. The mathematical part was made according to [1], [6] and [10]. Finally, some symptomatic items have been re-examined by using the 5-scale Likert score. For clarity for the readers in the following, we are describing briefly the Rasch analysis steps and the Likert method. So, according to the Rasch analysis steps, from the table R above we calculated initial students' ability β_n and item difficulty δ_i by using logit functions, [17], [18]

$$\delta_j = \ln\frac{1-P_{correct}(j)}{P_{correct}(j)} \qquad (1)$$

$$\beta_i = \ln \frac{P_{correct}(i)}{1-P_{correct(i)}} \qquad (2)$$

The probabilities $P_{correct}(i;j)$ are the respective ratios of a correct answer (the average by rows or columns respectively). Next, the difficulty is adjusted $\delta_i = \delta_i - <\delta>$ to shift the difficulty centered at zero. Based on the average values δ_i, β_j , the table of the probabilities estimate $P_e(i,j)$ for dichotomous variable are generated using the formula

$$P_e(i,j) \equiv P\left(x_{i,j} = 1 \middle| \beta, \delta\right) = \frac{\exp(\beta_j - \delta_i)}{1+\exp(\beta_j - \delta_i)} \qquad (3)$$

Next, initial R(i,j) elements are replaced by estimated values $P_e(i,j)$, and the process is performed iteratively until the summed squared residuals between expected probabilities and raw ones lies below a threshold. Detailed arguments can be found in [17], [19]. Similarly, for the n-scale Likert the probability that the person (n) would choose the category (h) for the item (i) is

$$P(x_{ni} = \mathrm{h}|\beta, \delta) = \frac{\exp \sum_{j=0}^{h}(\beta_n - \delta_{ij})}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k}(\beta_n - \delta_{ij})} \qquad (4)$$

where δ_{ij} represent the adjusted difficulty for the item (i) and depends on the measurement model. The final student ability β_j^{final} and problem difficulty δ_j^{final} are used for analysis and discussion. Particularly, the δ_j^{final} obtained for the standard FCI item would guide us on the problematic precipitation and features, the natures of obstacles that students meet when answering conceptual questions, typical and characteristic common sense beliefs used instead of conceptual knowledge and many other learning issues. The high variance cases are identified as misfitted cases called outfit and infit, which are analyzed too. Note that the reasons for discrepancies of the fit in Rasch model are subjects with high ability answering an easy question incorrectly and subjects with poor ability answering a hard question correctly by guessing. Therefore, we have additional information to analyze.

II. THE ASSESSMENT OF CONCEPT INVENTORY SCORES AND PRELIMINARY REMARKS

In our investigation herein we have used standard FCI test and other simplified or ad-hoc CI prototype. The Force Concept inventory (FCI) test has been used largely as a measure of student understanding of introductory mechanics [5], [14], [16]. It contains 30 typical multiple-choice items in mechanics [2], that are based on the 'six dimensions of the Newtonian concepts' and corresponding common-sense beliefs: (a) kinematics difficulties and vector issues, (b) impetus or intrinsic force needed to keep things moving, (c) active force misconception, (d) the prevalence against the third law, (e) dominances of the influences instead of superposition principles and (f) mixed obstacle influences in motion. Additionally, simplified FCI tests (SFCI) are used [10]. In our SFCI version, we regrouped the alternatives to facilitate the answering, and reformulated some items to enhance the readability of the text. This survey is conducted on four groups of 40-50 high school students from four main cities of the country, that also have had mechanics lectures by online system as result of pandemic closure. The hereto average FCI scores for four groups was found in the range $[8.43 - 9.75]$, that is lower compared to the regularly procedural exams and official results on maturity examination. We observed the highest CI-score value (14.04) belongs to Tirana city' students, who also have reported that they "*have had laboratory work and demonstration during physics lectures*". On the other side, the averaged FCI score for the mixed sample of 213 participants is estimated at 9.13. We observe that students who declared *"rarely or missing demonstration in the physics classes*" have been ranked as of low ability to answer the FCI test. From a point wise consideration, we noticed that *commonsense of conflict* is used mostly when interpreting *action/reaction* issues, and the most recurrent incorrect answer was observed in the categories of the *active force* and *concatenation influences*. We hypothesized up here that contextual reading was likely to be a cause factor responsible for low FCI score. For a better estimation of contextual understanding effect, we have realized a SFCI test. Note that it has not been possible to conduct both tests to same students (only a few students have provided 60 items answered!), but we managed to do them in same schools, hoping that the conclusions have remained intact. The SFCI score for each county has resulted in the range $[12.34 - 14.04]$, and totaling scores was found $SFCI_{score} = 13.67$. The relative improvement is estimated like the normalized FCI gain discussed in [14]

$$C = \frac{\%Scores_{SFCI} - \%Scores_{FCI}}{100 - \%Scores_{FCI}} = \frac{\left(\frac{13.67}{30} - \frac{9.13}{30}\right) * 100}{100 - \frac{9.13}{30} * 100} = 0.22. \quad (5)$$

This result suggested that students have found reading difficulty when answering the original FCI test, and we classify this as of contextual nature. This phenomenon is customary if physics lectures are introduced in a "mathematical fashion", without demonstration or laboratory support. Under those circumstances, students evoke their common sense to portray a physics situation needed to answer an item quest, resulting in wrong answers. Next, we remarked that some questions that directly involve the second Newton Law, have resulted with more miss fitted occurrences for FCI and SFCI tests. It contradicts our prior belief that such items must have clearer responses (not necessarily correct) because the Second Law of Newton is the most frequently mentioned topic taught in physics classes. We hypothesize that when Second Law is instructed narratively and without sufficient laboratory demonstration, students could be confused easily, and contextual shortcoming become important. The lack of laboratory works has been by nature the most apparent shortcomings during pandemic closure. It darkened the solid and natural perception of cause-consequence relationship that constitute Second Law basics. Accordingly, it is manifested as contextual misconception and ambiguity on answering relevant questions of standard FCI test. When simplifying and providing explanation for the context as we did in SFCI test, those effects become less influential, and the CI score get improved.

III. EVIDENCE FOR THE GAIN ON CONCEPTUAL KNOWLEDGE'S AFTER UNIVERSITY PHYSICS COURSES

Herein, we considered the improvement in the CI scores after 'a full reset' of the student knowledge inventory during university study. The participants in this survey were from first-year students of chemistry, applied mathematics, and informatics branches, at the Faculty of the Natural Sciences, University of Tirana. This test was realized in two distinct phases, at beginning and at the end of physics course. The first phase took place in November 2021, and the second phase at the end of March 2022. The results are displayed in Table 2. Here we obtained $FCI_{Score} = 10.55 \pm 2.8$ before the course and $FCI_{Score} = 14.35 \pm 2.3$ after it. We observe that after the academic course, the overall CI level was not satisfactory. It's worth to remark that the CI scores obtained in formal exams of general physics which are mostly procedural, have been remarkable better than the CI score measured hereto. Meanwhile, the advance for conceptual knowledge has remained problematic. The CI gain is evaluated

$$g = \frac{\%Scores_{AfterCourses} - \%ScoreBeforeCourse}{100 - \%Scores_{AfterCourses}} = \frac{\left(\frac{14.35}{30} - \frac{10.55}{30}\right) * 100}{100 - \frac{14.35}{30} * 100} = 24.3\% \quad (6)$$

Note that those students participating in this survey belongs also to the compulsory on-line education during pandemic closure. It seems that the stock of deficiencies on conceptual knowledge being created earlier, persisted in the subsequent stage of the studies.

Table 1 Results of double FCI test

	Branch	Chemistry	Applied Mathematics	Computing Science	Total
Nr. Students		20	37	64	121
CI	Before course	10.61	10.53	10.54	10.54
	After course	16.01	14.05	14.01	14.35
	Lab. in syllabus	Yes	No	No	

It resulted that chemistry branch' students have had the best gain of the FCI scores, estimated at 23.5%. Also, the calculated difficulty parameter δ for items related to the Second Law (items involving *active forces* in FCI test), have resulted smaller for this group. Noticing that regular course of physics for this branch includes laboratory works, this result indirectly indicates that the impairment on conceptual inventory observed by our survey should be related to the insufficient laboratory work and demonstration in support pf the physics learning inflicted by the online learning during pandemic. For completeness of the analysis, we have considered two more CI tests, a motion CI test based on general kinematics, and an Electromagnetism CI test. The EMCI test is developed by contemplating arguments of references [9], [10], [14], [17] and by exploiting our teaching experience. The findings are shown in Table 3. The observed gain of CI scores has resulted in nearly the same degree for all CI tests.

Table 2 Concept Inventory Tests

CI	Before the course	Std	After the course	std	The gain	Outfit at 1.3
FCI	10.55	2.78	14.35	2.33	24.3%	5
MCI	10.25	2.86	15.35	2.45	34.8%	5
EMCI	13.45	2.81	16.27	3.25	20.5%	7

Again, it resulted that for students that have followed on-line learning during pandemic, the improvement of CI scores after physics course was unsatisfactory.

IV. A DISCUSSION OF THE FCI ITEM DIFFICULTIES IN THE FRAMEWORK OF CONTEXTUAL KNOWLEDGE ISSUES

Another interesting observation from the tests conducted so far consists in ambiguities observed when encircling correct alternatives. It resulted by naked eye that after selecting an alternative as the correct answer, students come back and checked another alternative. This indecisiveness must be clarified. It resulted also that this hesitancy has appeared mostly on items where contextual formulation is important. So, in the SFCI version of the test, we reduced the contextual conceptual argumentation requirements. Just for illustration, the item 18 (in standard FCI) is reformulated as follows: "*by remembering that the acceleration on the constant velocity motion is zero, mark the correct answer for the forces acting on the elevator in the figure. The velocity of the elevator is constant”*. In this formulation, students don’t need an advanced contextual thought, enough to remember “mathematically’ the law in the way it was lectured. Note that it prompts for more guessed answers which consist of the occurrence of the unexpected probability that a student with given ability level can solve an item of a higher difficulty than its calculated ability. Next, the number of misfitted cases varied among four main-cities groups from 34 - 46 at the level 1.5, and from 48-62 at the threshold level of 1.3 suggested in [18]. Considering that total number of matrix elements in R was 1380-1680, ($\#students x 30 items$), those values are not neglect able and so is the difference observed. We observed also that outfitted items of the SFCI test were reduced to 32-46 entries at the level

1.3. The decrease of items being answered in a fuzzy way when applying SFCI tests, supports the idea that less contextual complexity makes the understanding of the questions homogeny clearer in the population. It is not an unextend issues, but support the major concern that contextual difficulty prevails. The difficulties $\delta_{SFCI} \in [-0.9, 0.47]$ have lower margins than FCI test difficulties $\delta_{FCI} \in [-1.2, 1.9]$. It resulted that simplification of the context for some items and reduction of alternatives in the multiple-choice questionary has lessened the difficulty ranges. Those findings are considered as indicatory of the significant sensibility of giving a correct answering from the context. Also, the relatively lower guessing occurrence in the SFCI tests can be interpreted as exhibition of the contextual CI impairing. To clarify the importance of this element, we have used in the following, the Likert 5-scale measurement tool.

V. A LIKERT SCALE ANALYSIS FOR CONTEXTUAL CONCEPTUAL KNOWLEDGE

In this last survey, we have assumed that if a definition of a physics quantity is given from a casualty perspective, the answers should be based on the context given in the formulation. Usually, the causality standpoint is reinforced by demonstration and laboratory work, but more factors are involved as we discussed above. Hereby, we proposed a finer measurement by using the Likert 5-scale method, considering arguments provided in [15], [25], [20], [21] etc. The test consists in three questions evoking the dependency of the answer from contextual formulation. For example, the mechanics item is: "*Based on the experimental observations, the acceleration of the point mass m under the net force F is proportional to the force and inverse proportional to the mass. It resulted that formulation for the Second Law is $F = ma$*". The same approach is utilized for Ohm' law. So, the context is fixed and should guide students' answers: one measures the consequence by adjusting the cause, and therefore, the correct contextual answers should be "$a = \frac{F}{M}$; $I = \frac{U}{R}$". Third item considers more complicated reasoning for the matters' waves. Question 3 is "*by replacing relativist energy $E = mc^2$ and taking $P = \frac{mc^2}{c} = \frac{hv}{c} = \frac{h}{\lambda}$ it resulted that $P = \frac{h}{\lambda}$ represents a real and a measurable physical quantity known as photon momentum. Therefore $\lambda = \frac{h}{P}$ defines a measurable and real physical quantity as a direct mathematical consequence* ". In this test have participated 175 students. We noted that this survey has been conducted separately from other surveys analyzed in this study, but results are useful for iterating step by step findings herein. It took place in March 2021. The results are displayed in the Table 4.

Table 3. Likert values

	Item 1	Item 2	Item 3
Likert Value	2.12	3.08	3.55

Std	0.58	1.23	1.55
Likert Value	3.49	2.52	1.89
Std	0.86	1.35	1.59

From the Table 5, it is seen that the average Likert value answer of mechanics quest lies within the limit of the interval null, indicating a high ambiguity in answering. They agree with the contextual formulation of the second law $a = \frac{F}{m}$ $(3.49 \in [3.4,4.2] \rightarrow agree)$ $(2.2 \in [1.8,2.6] \rightarrow disagree)$ and they have no decision on the contextual Ohm law. It revealed the problematic conceptual knowledge in physics for high school's students that learned online the full courses. The most complicated item, question 3, reveals another problem: the extensive tendency for mathematical reasoning when dealing with a physics conceptual question. Also, the deviance (standard deviation) is vast for all cases analyzed. We concluded that in this survey, many respondents answered the questions by contemplating visual remembrance of the relationship ($F = ma$ is widely used in lectures and literature) or by employing apparent mathematical symmetries instead of physical reasoning. The causality context is ignored in those cases. It reinforces the above findings for the reduction of the conceptual reasoning as direct effect of the enforced online learning during pandemic closure and other factors that have reduced the weight of laboratory work, demonstration etc.

VI. DISCUSSIONS AND CONCLUSIONS

The physics CI_{scores} obtained from students whose have had their high school study during the pandemic period 2019-2021, belongs to the "low and insufficient knowledge in Newtonian mechanics", at the range 8-12 points from 30. Also, those results are significantly smaller than general knowledge scores obtained from regular exams. In general, the ability to solve the FCI test has resulted lower for students that reported missing or rare demonstrations in the physics classes during the period referred. The calculated difficulties as perceived from the students participating in the survey, have resulted hardened for FCI test items that are strongly conditioned form the practicing and concretization during learning process. We interpreted those findings as indicators that the most influential factor in this problematic evidence is the lack of demonstration and laboratory work that culminated during 2020-2022 pandemic closure. The Conceptual Inventory Scores for the bachelor students that followed physics course as by their syllabus, has improved 20%-38% after the course, and the best improvement is obtained for students that do have laboratory works in their program. This gain is estimated also as not satisfactory, which designate the persistence of conceptual knowledge deficiencies inherited from previous phase of study. By comparing the scores obtained in FCI and SFCI tests and reanalyzing the outcomes by a tinier measurement instrument, we concluded that contextual errors are the most frequent. The findings of this work can be considered as indicatory for negative effect of compulsory online learning on the conceptual physic knowledge of high school students. Also they supported the idea that observed failures is caused mostly from the temporal teaching's impossibility to develop live demonstration and other means of concretizing interactions that could help in correct transmitting of conceptual knowledge to the students. The lack of solid knowledge due to this shortcomings favors alternative "mathematically" thinking in the best opportunity, or worse, using commonsense and (wrong) beliefs to interpret a physical situation. Also, it explain the slightly better level on mathematical knowledge compared to the physical ones as observed herein too. Other potential factors may be present at have affect also students' knowledge in physics, but we do not individualize them through this short view and deeper analysis are needed. We clam that a through measure including teaching performance would shed more light, which remain for considering in the forthcoming works.

References

1. N.M.Crooks, M.W. Alibali. Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, 34 344–377.
2. D. Hestenes, M. Wells, M.G. Swackhamer. Force concept inventory. *Phys. Teach*. 30, 141
3. A. Savinainen, F. Scott. The Force Concept Inventory: a tool for monitoring student learning. *Phys. Educ.* 37 45.
4. R. Rahmawati, N. Rustaman, I.Hamidah, D. Rusdiana. The Development and Validation of Conceptual Knowledge Test to Evaluate Conceptual Knowledge of Physics Prospective Teachers on Electricity and Magnetism Topic. *Jurnal Pendidikan IPA Indonesia*, 7(4), 283-490.
5. R.K. Thornton. Using large-scale classroom research to study student conceptual learning in mechanics and to develop new approaches to learning in Microcomputer-Based Labs: Educational Research and Standards, 156, 89-114.
6. A. Savinainen, J. Viiri. The Force Concept Inventory as a Measure of Students Conceptual Coherence. *International Journal of Science and Mathematics Education*. 6(4),719-740
7. P. Luangrath, S. Pettersson, S. Benckert. On the Use of Two Versions of the Force Concept Inventory to Test Conceptual Understanding of Mechanics in Lao PDR. *Eurasia Journal of Mathematics, Science & Technology Education*, 7(2), 103-114
8. C.J. Anderson, Z. Li, J.K. Vermunt. Estimation of models in a Rasch family for polytomous items and multiple latent variables. Journal of Statistical Software, 20.
9. B. M. Notaros. Concept inventory assessment instruments for electromagnetics education. *IEEE Antennas and Propagation Society International Symposium*. 1(1),684-687
10. M.A. Nelson, R.L. Miller, R.A. Streveler, B.M. Olds, and M.R.Geist. Using concept inventories for formative assessment of conceptual learning: A case study from engineering. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, California April 7, 2006.
11. A. O'Shea, S. Breen and B. Jaworski. The Development of a Function Concept Inventory. *Int. J. Res. Undergrad. Math. Ed*. 2:279–296
12. F. Huriawati, N. Fitriani, J. Handhika. Force concept inventory (FCI) representation of high school students (SMA & MA), *J. Phys.: Theor. Appl*. 1(1), 29-34
13. F.M. Lord. Applications of Item Response Theory To Practical Testing Problems (1st ed.). Routledge. https://doi.org/10.4324/9780203056615

14. M. Planinic, M. Assessment of difficulties of some conceptual areas from electricity and magnetism using the conceptual survey of electricity and magnetism. *American Journal of Physics*, 74(12), 1143–1148.
15. J.Camparo, L.B. Camparo. The Analysis of Likert Scales Using State Multipoles: An Application of Quantum Methods to Behavioral Sciences Data. *Journal of Educational and Behavioral Statistics*, 38(1), 81–101. http://www.jstor.org/stable/23324435
16. M. Planinic, L. Ivanjek. A. Susac. Physical Rasch Model Based Analysis of the Force Concept Inventory. *Physical Review Special Topics - Physics Education Research* 6, 010103
17. R. Chabay, B. Sherwood. Brief Electricity and Magnetism Assessment. *Physical Review Special Topics - Physics Education Research*, 2(1), 7-13
18. https://www.real-statistics.com.
19. C.J Anderson, J.Verkuilen, B.L. Peyton, B. L. Modelling Polytomous Item Responses Using Simultaneously Estimated Multinomial Logistic Regression Models. *Journal of Educational and Behavioral Statistics*, 35(4), 422–452.
20. D. Bertram. Likert Scales: CPSC 681—Topic Report. Poincare, 1(11).
21. S.A. McLeod. Likert Scale Definition, Examples and Analysis. *Simply Psychology*.