THE EUROPEAN EDUCATIONAL RESEARCHER
DOI: 10.31757/euer.313

http://www.eu-er.com/

Development of Modern Physics Achievement Test: Validity and Reliability Study

Nuri Balta & Ali Eryılmaz

Middle East Technical University, Turkey

Abstract: In this study, we aimed to develop an instrument that could be used to measure students' achievement in modern physics at high school level in a valid and reliable manner. The study was carried out in fall 2013–2014 with a total of 304 students. In this study, expert opinions were obtained to determine the test's content validity. The reliability of the test was obtained via Cronbach's alpha, which produced reliability coefficients that fell within acceptable limits. Item analysis were conducted to eliminate improper items. Based on these findings, it could be concluded that the test is an instrument that produces valid and reliable measures, and that can be used to determine students' achievement in modern physics.

Keywords: *Modern physics test; Physics achievement test; Test validity and reliability*

Introduction

Modern Physics (special relativity): When the physics education literature was examined, it was observed that until today the number of studies done about students' understanding of relativity is quite small in number and their focus is mostly on Galilean relativity (Selçuk, 2010).

The special relativity concepts are so counterintuitive and contradicting with our daily understanding of space and time that physics students find it hard to learn relativity (Hewson, 1982; Scherr, 2007; Villani & Pacca 1987; Hosson, Kermen & Parizot 2010; Selçuk, 2010).

Previous studies about theory of special relativity are not very numerous and they showed that students fail in defining and using the concepts of theory and thus confuse most of its concepts (Hewson, 1982; Scherr, 2007; Villani & Pacca, 1987). The results of the examples listed below supports these findings.

First, Hosson, Kermen and Parizot (2010) in their study aimed at exploring prospective physics teachers' reasoning associated with the concepts of reference frame, time and event which form the framework of the classical kinematics and that of the relativistic kinematics. The research was conducted in France and 94 prospective physics teachers were surveyed by means of a questionnaire. The students responded to eight multiple choice questions including a request for justification. Their results showed that students show a deep lack of understanding of both concepts of reference frame and event.

Second, the aim of the study conducted by Dimitriadi and Halkia (2012) was to investigate students' learning processes towards the two axioms of the theory of special relativity (the principle of relativity and the invariance of the speed of light) and the consequences of the two axioms. They developed a teaching and learning sequence consisting of five sessions after analysing the physics college textbooks, reviewing the relevant bibliography and conducting a pilot study. To collect the data, they used experimental interviews. Their sample consisted of 40 10th grade students. They collected the data by interviews, as well as by two open-ended questionnaires filled out by each student, one before and the other after teaching

theory of special relativity. Their results showed that upper secondary education students were able to cope with the basic ideas of the theory of special relativity, however they found that the conceptions; (a) there is an absolute frame of reference, (b) objects have fixed properties and (c) the way events happen is independent of what the observers perceive were difficult for students to understand.

Third, Selçuk (2010) investigated the pre-service teachers' understanding of and difficulties with some core concepts in the special theory of relativity. The 185 participants were from the Departments of Physics Education and Elementary Science Education at Dokuz Eylul University. She used both quantitative and qualitative research methods in her study. She applied a paper-and-pencil questionnaire including four questions and conducted in-depth interviews with the participant teachers after the instruction of related modern physics topics. Pre-service teachers' understanding of and difficulties with core elements of special relativity such as time, length, mass and density were tested. Teachers' specific and considerable difficulties with proper time, time dilation, proper length, mass and relativistic density concepts were among the results of her study. After examining the related literature, she summarized that no matter from which academic level (i.e. from secondary to university) the students are obviously have difficulties in understanding and comprehending special relativity subjects.

Even though they didn't analyse the prospective teachers' understanding of special relativity theory, Yıldız (2012) and, Hosson, Kermen and Parizot (2010) found in common that prospective physics teachers have a deep lack of understanding of concepts associated with special relativity.

A brief search of the literature revealed that the number of publication on modern physics (special relativity) is limited. Moreover, due to counterintuitive nature of the concepts students find it hard to learn and to understand the deep implications of the theory. There exists a need to contribute to the literature by developing a test that can be used to measure students' achievement in modern physics.

Sample

There were two samples in this study. The pilot study was done with 42 students from a private school in Ankara. The main sample of this study (n=306) was composed from six schools, all tenth grade students, from Anatolian high schools located at Demetevler quarter of Yenimahalle district and city centres of Mamak and Altındağ districts of Ankara. Subjects from six schools out of 24 in these districts have almost the same characteristic in terms of prior achievement and socioeconomic status. The ages of students in both pilot and main group ranged from 15 to 17 and while main group students were all at 10^{th} grade, pilot group students were from 11^{th} and 12^{th} grades. 43% of main group and 45% of the pilot group were female.

Development of the Achievement Test

In order to measure students' academic achievement in tenth grade in Modern Physics Unit (MPU), an achievement test was developed by the researcher. This unit was chosen because of the fact that it was recently added to the national tenth grade physics curriculum (2007) and teachers had experienced problems in teaching this unit (Eryılmaz, 2012).

Before starting construction of the test, objectives of the unit, which were determined and declared by the Ministry of Education, were examined. A table of test specification that represents the content of the

tenth grade MPU was prepared. Five objectives were specified for this unit in the curriculum. However, since one objective was not taught in schools which have weekly two physics course hours, this objective was not represented by any questions. The weight of each objective was determined according to the time allocated to teach each of them. Further, the questions' difficulty level was determined according to the specifications of the curriculum and thereby was determined according to Bloom's revised taxonomy (Krathwohl, 2002).

In developing the achievement test, following issues were taken into consideration. First, one lecture hour is 40 minutes in high schools and the test must be finished in a class hour. Second, the curriculum requires context based questions. Third, context based questions generally have long stems and this property usually bore students. On the other hand, finding or writing questions on tenth grade MPU is a little bit troublesome. This unit has been added to the Turkish physics curriculum since 2008 and textbooks published by various companies do not include desired questions that reflect the aim and the requirements of the curriculum. That's why except two, all 32 questions on the final version of the Modern Physics Unit Achievement Test (MPUAT) were written by the researchers.

The development phase of the achievement test approximately took three months. Initially, two successive rough versions of the achievement test were prepared by the researcher, then the first version was developed by the researcher. Afterwards, the first version was checked by the experts and upon their requests second version was generated, then the validity of the second version was confirmed by the second expert review process and finally, the final version constituted after item analysis.

At the very beginning, the researcher prepared the first rough version of the achievement test. He prepared 26 questions by referencing the tenth grade physics course books. The researcher and one of his colleague discussed all questions one by one. However, questions that were problematic in terms of objectives (5 items) were removed, that were not comprehensible (10 items) were revised and that included conceptual errors (6 items) also were removed. Thus, 11 questions were deleted from the first rough version of the achievement test. Then the researcher examined books such as Conceptual physics (Hewitt, 2006), The Physics for Everyday Phenomena (Griffith, 2001) and Physics for Scientist and Engineers (Serway, 2004). The researcher than read the tenth grade MPU related topics from these books and he prepared a second rough (15 questions were from the previous version) test consisting 30 questions. None of the added questions were directly taken from the books. They all were written by the researcher after studying the related chapters from the aforementioned books. The researcher and one of his colleague again did long discussions on almost all questions to prepare the first version of the achievement test. They revised some questions (5 items), changed the structure of some questions (4 items) and removed some questions (9 items) from the second rough version. Then, the researcher wrote new items and added some new multiple choice questions (6 items), added some new true-false (7 items) questions, added one matching question and finally he generated the first version of the achievement test.

To ensure face and content validity expert views were asked from four experienced teachers that were serviced full-time in private high schools. Moreover, expert views of four research assistants in the department of Secondary Science and Mathematics Education were taken. Furthermore, the views of

two physics course book writers of a special company were also taken. Together with the achievement test, (1) the tenth grade modern physics unit curriculum, (2) the table of test specification, and (3) the expert opinion form for 10th grade modern physics test was sent to all these experts.

Upon the request of experts, major modifications made in the first version of MPUAT. Some of the questions (15) were removed and replaced with new questions. The reason for removing so many questions was not that they were wrong questions; upon the request of experts they were replaced with true-false and matching questions. Actually most of these questions changed their structure from multiple choices to true-false, matching or open ended. Moreover, distractors of some questions were revised (3 items) and some of the questions (5 items) were reworded upon the request of the experts. Two questions were declared by the experts that they were not related to the objectives of the tenth grade MPU, that's why they were replaced with new questions. Moreover, 34th question was a matching question it was both reworded and its structure was redesigned. Furthermore, upon their requests a sixth distractor (I don't know the answer) was added to all questions. Thus, the initial extensive expert views leaded to major revision in the first version of MPUAT.

After the expert views, since too many changes were made on the first version of MPUAT, one more expert view became necessary for the second version. The second version of the MPUAT was examined by four experts, three of which were the same experts who examined the first version of MPUT and one was a new expert. Except some minor changes all experts were agreed on the face and content validity of the second version of the MPUAT. Additionally, they were asked to generate an answer key for MPUAT questions. The answer key generated by these experts was same as that of the researcher. Thus, the validity of the second version of MPUAT was confirmed by the experts and thus the achievement test was developed after two stage expert view process. After the final revision, the second version of the MPUAT had 32 items: 6 true-false, 6 matching, 18 multiple-choice, and 2 open-ended items.

Prior to pilot study of second version of MPUAT, two students at different achievement levels in physics read the questions loudly. The researcher listened to the students and tried to catch the points where students have difficulties in understanding. However, the researcher didn't saw any problems; moreover, the students didn't report any misunderstandings. These two students were from a private school and they were at 12th and 11th grades. The former was a high achiever and the later was a normal student. Both were taught MPU previously. Each of these practices lasted approximately 45 minutes.

As a pilot study, the MPUAT (second version) was administered to 42 11th and 12th grade students from a private high science school in Ankara. These are the students who have learned the MPU in 10th grade and among them the 12th grade students have re-taught this unit in university preparation courses. There were several reasons to choose such a school and the combination of 11th-12th grade students for pilot study. First of all, the test had to be applied to a sample which already has mastered MPU. However, there weren't such a sample. Secondly, since MPU was newly added to the curriculum many teachers had superficially taught this unit. However, private schools relatively teach better and educate their teachers in the case of any changes of curriculum. Thirdly, MPU is not interrelated to

remaining 10th grade and 11th grade units, that's why students who participated to the pilot study didn't find any chance to repeat at least some of the concepts of this unit. Fourthly, the science high schools generally consist of students at high achievement level.

Item difficulty and item discrimination were conducted with data gathered from these students. Table 1 shows the item analysis results.

Table 1

Item Analysis Results of the MPUAT-S for Pilot Study

Item #	Difficulty	Discrimination	Item #	Difficulty	Discrimination
1	0.78	0.27	17	0.31	0.55
2	0.5	**-0.1**	18	0.37	0.45
3	0.31	0.45	19	0.33	0.27
4	0.64	0.36	20	0.48	0.73
5	0.55	0.64	**21**	0.24	**0.09**
6	0.48	**0.09**	22	0.52	0.73
7	0.56	**0**	23	0.26	0.27
8	0.83	**0**	24	0.43	0.64
9	0.4	0.45	25	0.81	0.27
10	0.62	0.36	26	0.48	0.45
11	0.64	**0.18**	**27**	0.88	**0.18**
12	0.76	**0.09**	28	0.27	0.55
13	0.26	0.36	**29**	0.18	**0.18**
14	0.12	**0.09**	**30**	0.81	**-0.3**
15	0.76	0.45	**31**	Not enough correct answers were given to the two open ended questions	
16	0.54	**0.09**	**32**		

# of Items	30	Variance	0.79	Kurtosis	-0.29
# of Examinees	42	Std. Dev.	4.56	Alpha	0.72
Mean	14.98	Skew	0.35	Mean item difficulty	0.50
				Mean item discrimination	.30

*Bold are the questions having improper item discrimination indices. In other word those have indices smaller than 0.19.

It can be seen from Table 1 that the item discrimination indices of items are in the range of -0.30 to 0.73. The items that have values under 0.19 should be removed or completely revised. Moreover, the items that have values between 0.20

and 0.29 can be checked for modification (Crocker & Algina, 1986, p. 315). Table 1 indicates that items 2, 6, 7, 8, 11, 12, 14, 16, 21, 27, 29 and 30 should be removed or completely revised and items 1, 19, 23 and 25 can be modified. Since removing so many items would affect the validity of the test, the item correction or reformation was postponed to the data gathered from the main group. Moreover, the average item difficulty for test items was 0.50 and the internal reliability coefficients for the test were found as 0.72. Except the changes made on the distractors of 13 questions, the MPUAT was not revised with respect to the results of the pilot study. Since none of the students correctly answered the two open ended questions (31st and 32nd items) they were not included into item analysis. According to these analyses, a total of 12 questions (Table 1) were problematic in terms of item discrimination they were considered to be removed from the test. However, because of several reasons all items kept their places in the test. Firstly, average score on the test was medium (an average of 15 correct answers), that's why it was taught that the selected sample might not be favorable. Secondly, the number of questions was appropriate; during the application of the test it was seen that students could easily find enough time to answer all questions, that's why there was no need to decrease the number of items. Thirdly, a sample, who has just learnt MPU, could give more accurate results. Due to all these reasons the items that were going to be removed was postponed. However, distractor analysis was performed on the data gathered from this sample. According to item analysis conducted via ITEMAN program, some mild to moderate modifications, based on alternative statistics, were made on the choices of 13 questions of second version of MPUAT.

In items 13, 15, 17, 18, 19, 23, 25, 26 and 28, since one distracter of each were selected with a low rate, the alternatives were rearranged. In question 22, since one distracter was selected with a low rate the figure of this distracter was redrawn. In question 14 and 21 since one distracter of each were selected with a high rate it was replaced with a new alternative. In question 30, two distracters were selected with a low rate. The alternatives of this question were revised. Thus, the final version of the MPUAT was constituted. Except two (23rd and 27th questions), all other questions were written by the researcher.

Each question in the MPUAT has an extra option "I don't know / I can't do". If students had chosen this option, in the scoring process it was coded as "0". In this way, we can see if unanswered questions are missing or students read the questions and do not know the answer. During the administration of the tests which was done by the researcher, the students were encouraged to circle the "I don't know / I can't do" alternative in the case of having no idea about the answer of the question. After the test was applied to 304 students, it was checked and seen that all students have either chosen one of five alternatives or have chosen the sixth alternative ("I don't know / I can't do"). In other words, there were no unmarked items in the answers.

The questions that were excluded from the analysis were determined according to the item analysis made on the data gathered from test scores of the main group. Totally, the data of 116 subjects were used for this analysis. Since (1) the main group students had newly learned the MPU, (2) their teachers participated to a PD program related to modern physics unit (Balta, & Eryılmaz, 2019), (3) the number of subjects were good enough, item analysis was conducted on data gathered from this

sample. As in the case of pilot study, the two open ended questions' results were undesirable, that is almost all students in the main group either gave wrong answers or didn't provide any answers. In other words, the mean of the 31st question was 0.11 and that of 32nd question was only 0.01. Consequently, they were excluded from item analysis.

Item analysis results of the MPUAT for test scores of main group subjects is given in Table 2.

Table 2

Item Analysis Results of the MPUAT for Main Group Subjects

Item #	Difficulty	Discrimination	Item #	Difficulty	Discrimination
1	0.62	0.35	17	0.33	0.45
2	0.16	**-0.10**	18	0.20	0.26
3	0.58	0.65	**19**	0.15	**0.03**
4	0.34	0.35	20	0.14	0.39
5	0.43	0.32	**21**	0.41	**0.10**
6	0.55	0.26	22	0.17	0.29
7	0.48	0.39	**23**	0.06	**0.10**
8	0.66	0.29	**24**	0.19	**0.06**
9	0.42	0.61	25	0.41	0.39
10	0.50	0.61	26	0.27	0.55
11	0.43	0.58	27	0.55	0.61
12	0.77	0.35	**28**	0.05	**0.06**
13	0.21	0.35	29	0.28	0.42
14	0.37	0.52	**30**	0.28	**0.06**
15	0.41	0.55	**31**	Not enough correct answers were given to the two open ended questions	
16	0.32	0.19	**32**		

# of Items	30	Variance	18.41	Kurtosis	0.21
# of Examinees	116	Std. Dev.	4.29	Alpha	0.70
Mean	10.76	Skew	0.64	Mean item difficulty	0.36
				Mean item discrimination	.34

Table 2 indicates that the item discrimination indices are in the range of -0.10 to 0.65. Moreover, the average item difficulty for test items is 0.36 which means only 36% of the participants answered test items correctly. According to results of last item analysis, 9 of the items were excluded. Items 2, 19, 21, 23, 24, 28 and 30 were removed because their item discrimination indices were smaller than 0.19.

In addition, the item 6 has a discrimination index of 0.26 and item 16 has an index of 0.19. These two questions were kept in the test in order to keep the percentages of the objectives in table of test specification. Moreover, they were checked and it was seen that there was no need for the changes. The two open ended questions (31 and 32) were removed because, as mentioned above, students either didn't provide answers to these questions or mostly gave wrong answers.

Removing these items neither affected the content validity of the MPUAT nor did not decreased the number of items too much. There were four objectives assessed with the achievement test. Before removing these items each objective, in average, was assessed by eight items and after removing these items the objectives, in average, were assessed by 5.75 items. Moreover, the achievement test includes questions at two levels (understanding and analysis) of Bloom 's taxonomy and removing these items didn't affect the percentages of these levels of Bloom's taxonomy. Furthermore, it seemed that totally items were so difficult for the students (0.36). This could force the students to guess answers even they did not really have any idea about them. Thus, the item discrimination indices could be affected by guessing. Once 9 questions were not included into analysis MPUAT's final version remained with 23 items: 5 true-false, 6 matching, and 12 multiple-choice items. The internal reliability coefficients for the data collected from main group was found as 0.70. After the extraction of the 9 items those do not work properly, the reliability of the test rose to 0.75. This value indicates high-medium reliability. This value could be because of guessing and unconscious answers of the students who faced with many concepts in the MPU for the first time. Since most of the items extracted from the test were difficult questions, average item difficulty decreased to 0.41 after extraction.

Finally, all test items (true/false, matching and multiple-choice) are coded as "0" for wrong and 'I don't know' answers, and "1" for correct answers. Each question was one point and for the lastly remaining 23 items subjects could have achievement scores ranged from 0 to 23. Higher scores indicate higher achievement level and lower scores indicate lower achievement level. The average completion time for the MPUAT was 40 minutes.

Discussion and Conclusion

The purpose of this study was to develop an instrument that could validly and reliably measure students' achievement in modern physics at high school level. Several versions of the test were developed and a two stage expert opinion process was conducted to ensure the scale's content and face validity. An achievement test of 23 items related modern physics was developed.

An item analysis was carried out in order to determine the how well the MPUAT items predicted the total score, as well as the items' levels of distinctiveness. These findings suggest that all of the items are discriminatory. The findings obtained in this research via statistical analyses, with the ultimate aim of examining the psychometric properties of the measures presented in the MPUAT, demonstrate that the scale can be used an instrument that produces valid and reliable measures to determine students' achievement in modern physics.

To our best knowledge, the literature shows that there is no scale that can be used to measure students' achievement in modern physics. Therefore, it is believed that this will significantly contribute to the relevant literature.

In addition to the strengths listed above, the present study has some limitations. These limitations bring about a number of suggestions for further research. First, this scale development study was carried out only in Ankara. Thus, studies should be carried out with different samples, as this is important for the replication of the validity and reliability analyses. Another suggestion for further research includes analyzing the concurrent validity of the MPUAT's measures. Future studies might investigate the kind of relationship that exists between MPUAT and other relevant measures.

References

Balta, N., & Eryılmaz, A. (2019). The effect of the 'teacher-led PD for teachers' professional development program on students' achievement: an experimental study. *Teacher Development, 23*(5), 588-608.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Holt, Rinehart and Winston INC.

Dimitriadi, K., & Halkia, K. (2012). Secondary Students' Understanding of Basic Ideas of Special Relativity. *International Journal of Science Education*. 1–18, First Article.

Eryılmaz, A. (2012). *Personal communication*, November, 29.

Griffith, W. T. (2001). *The Physics for Everyday Phenomena: A conceptual introduction to physics*, MCGraw Campanies.

Hewitt, P. G. (2006). *Conceptual physics*, Ninth edition, Pearson.

Hewson, P. W. (1982). A case study of conceptual change in special relativity: the influence of prior knowledge in learning. *International Journal of Science Education*. *4*, 61–78.

Hosson, C., Kermen, I., & Parizot, E. (2010). Exploring students' understanding of reference frames and time in Galilean and special relativity. *European journal of physics*, *31*, 1527–1538.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212-218.

Scherr, R. (2007). Modeling student thinking: an example from special relativity. *American Journal of Physics. 75*, 272–80.

Selçuk, G. S. (2010). Addressing pre-service teachers' understandings and difficulties with some core concepts in the special theory of relativity. *European journal of physics*, *32*, 1–13.

Serway, R. A., & Jewett, J. W. (2004). Physics for Scientists and Engineers 6th edn (Thomson Brooks/Cole)

Villani, A. & Pacca, J. L. A. (1987). Students' spontaneous ideas about the speed of light. *International Journal of Science Education, 9* 55–66.

Yıldız, A. (2012). Prospective Teachers' Comprehension Levels of Special Relativity Theory and the Effect of Writing for Learning on Achievement. *Australian Journal of Teacher Education. 37*(12), 15-28.

Corresponding Author Contact Information:

Author name: Nuri Balta

University, Country: Middle East Technical University, Turkey

Please Cite: Balta, N., & Eryılmaz, A. (2020). Development of Modern Physics Achievement Test: Validity and Reliability Study. *The European Educational Researcher, 3* (1), 29-38. DOI: 10.31757/euer.313

Received: August 27, 2019 ▪ *Accepted: February 01, 2020*