SOCIAL MENTALITY AND RESEARCHER THINKERS JOURNAL Doı: http://dx.doi.org/10.31576/smryj.563
RESEARCH ARTICLE SmartJournal 2020; 6(33):1168-1175
Arrival : 25/05/2020
Published : 20/07/2020

THE DEVELOPMENT OF ACADEMIC ACHIEVEMENT TEST FOR THE SCIENCE DISCIPLINE WITHIN WEDO 2.0 ROBOTIC SET

Wedo 2.0 Robotik Seti İle Fen Bilimleri Dersine Yönelik Akademik Başarı Testinin Geliştirilmesi

Reference: Bahçeci, F. & Üşengül, L. (2020). "The Development Of Academic Achievement Test For The Science Discipline Within Wedo 2.0 Robotic Set", International Social Mentality and Researcher Thinkers Journal, (Issn:2630-631X) 6(33): 1168-1175.

Dr. Öğrt. Üyesi Ferhat BAHÇECİ
Fırat Üniversitesi, Eğitim Bilimleri Fakültesi, Elazığ, Türkiye
ORCID: 0000-0001-6363-4121

Leyla ÜŞENGÜL
Fırat Üniversitesi,Eğitim Bilimleri Entitisü Yüksek Lisans Öğrencisi, Elazığ, Türkiye
ORCID: 0000-0002-1183-6888

ABSTRACT

With the changes in the education system for integrating 21st century skills, robotics and coding education have become important. LEGO sets are one of the basic materials to be used in robotic education. WeDo 2.0 robotic set aims to help students learn interdisciplinary information. Although robotics and coding education has a very important role today, there are only few measurement tools giving information about the achievement level. The purpose of this study is to develop an achievement test whose validity and reliability is provided for science lesson within robotics and coding education by using WeDo 2.0 set. The test was applied to 111 students studying in 6th, 7th and 8th grades. The test which initially consisted of 35 items was evaluated in terms of its appropriateness to the field and students' levels. However, as a result of expert opinion to ensure the validity and comprehensibility of the test four items were excluded from the test. Item analysis was performed on the questions. The difficulty values of the items in the final form of the test were between .17 and .91. The mean difficulty index of the test was found as 0.66. In terms of the reliability analysis of the test, KR-20 value was found to be .889. After the analysis, the test was structured as 24 item. As a consequence, the test was considered as a valid and reliable tool. An achievement test with high validity and reliability was prepared in order to measure target gains for science lesson within WeDo 2.0 robotic set.

Keywords: Robotics, coding, achievement test, science education, test development.

ÖZET

21. yüzyıl becelerileri için eğitim sistemindeki değişimler ile robotik ve kodlama eğitimi önemli bir yer edinmiştir. Robotik eğitimindeki materyallerin temelinde LEGO setleri bulunmaktadır. WEDO 2.0 robotik seti ile öğrencilerin disiplinlerarası bilgiler öğrenmeleri amaçlanmaktadır. Günümüzde robotik ve kodlama eğitimi önemli bir yere sahip olmasına rağmen başarı düzeyiyle ilgili bilgi verecek ölçme aracı oldukça sınırlıdır. Çalışmamızda; Wedo 2.0 setiyle verilen robotik ve kodlama eğitiminin, fen bilimleri dersine yönelik geçerliği ve güvenirliğini sağlama amacıyla başarı testi geliştirilecektir. Bu test 6. 7. ve 8. Sınıftaki 111 öğrenciye uygulanmıştır. 35 maddeyle oluşturulan test, geçerliği sağlamak amacıyla uzman görüşü alınıp anlaşılabilirliği, alana ve öğrenci seviyesine uygunluğu, açılarından değerlendirilmiş ve dört madde testten çıkarılmıştır. Test sorularına madde analizi yapılmıştır. Testin son formundaki maddelerin güçlük değerleri .17 ile .91 arasında değer almış, ortalama güçlük indeksi 0.66 bulunmuştur. Testin güvenirlik analizi KR-20 değeri ,889 bulunmuş ve yüksek güvenilirlikte olduğu saptanmıştır. Geçerlik ve güvenirlik analizlerinden sonra test 24 madde olarak yapılandırılmıştır. Testin geçerli ve güvenilir bir ölçme aracı olduğu belirlenmiş ve Wedo 2.0 robotik setiyle, fen bilimlerine yönelik kazanımların ölçülmesi için uygun bir başarı testi geliştirilmiştir.

Anahtar Kelimeler: Robotik, kodlama, başarı testi, fen eğitimi, test geliştirme.

1. INTRODUCTION

Until the recent past, technology was considered as tablet, phone, computer, electronic tools and equipment. However, nowadays, technology is considered as a wide frame as robots, coding, artificial intelligence, and fuzzy logic. Technology can be regarded as producing solution for problems and easing human life by the materials developed by integrating the concepts from different disciplines (Çepni, 2005). Technology is a process in which materials and systems are developed for people's needs (MEB, 2006).

Human life is shaped with phenomenon, case, process and technological materials related to sciences. Countries where manpower in the field of science is not enough are seen to be more dependent on other countries technologically (Çilenti, 1985).

Technology-supported science education is seen to be better for students to comprehend the concepts in terms of providing the learning by experiencing through putting the students into the center (Goldworthy, 2000; Jimoyiannis and Komis, 2001).

In developed countries, individuals are expected to have the skills of effective use of information technologies, critical thinking, problem solving, tendency to team works and taking responsibility. It is seen that the robotics and science activities are given to the students in an integrated structure in order to gain these skills. Although robotics education dates back to 1990s, it is seen just to be used recently in education field in Turkey. It was observed that robotic designs and team works in science education provided concrete learning on the students and had positive effects on their technology usage levels (Costa and Fernandes, 2004; Cited by Koç-Şenol, 2012).

The use of robotic technologies in science education has become very easy. Today, the development of the robot market and the increase of competition will increase the use of robots. Although there are more than one robot types in the educational robot field, Lego products are more preferred. Both the software and the lesson contents as well as the materials attract more attention among students and therefore, they are selected. Lego Education educational robot sets are available for students in all ages and levels such as Wedo, Wedo 2.0, NXT, EV3 Mindstorms.

Even though various educational robotic sets are used nowadays, they have some deficiencies in measurement. It is also observed that the effects of robot-based educations on different disciplines are examined in terms of variables such as attitude and motivation but there is no complete measurement tool or it cannot be applied related to the contents of robotic kits. After determining this deficiency, it was aimed to develop an achievement test in order to measure the target gains related to science contents of Wedo 2.0 robotic set.

In today's education system, measurement and evaluation are an integral part of the process. The measurement tools have an important place in determining the readiness of the students and their misconceptions in terms of affective and psychomotor aspects. At the beginning of developing process of measurement tool, the scope and purpose of the test should be clarified (Özçelik, 2010). Once the purpose and scope of the test are determined, the behaviors or educational outcomes that will confirm the learning process should be clearly determined.

It is very important to investigate the effects of interdisciplinary educations for 21st century skills and to develop different applications according to the results. In order to determine the learning levels of the students receiving robotic-assisted science education, it was aimed to develop an achievement test in accordance with the lesson contents within Wedo 2.0 robotic set.

2. METHOD

2.1. Study group

This study was conducted with 111 students studying in 6^{th}, 7^{th}, and 8^{th} grades in Elazığ Bahçeşehir College in 2017-2018 academic year.

2.2. Data collection tools

As data collection tool, a multiple-choice achievement test prepared with 31 items prepared in accordance with Wedo 2.0 robotic set and the contents of science lesson was used. The first step of the achievement test was to prepare a question pool with 35 items. Opinions from field experts including four from science education and one from educational sciences were obtained for the questions in the pool. The test, initially prepared with 35 items was evaluated in terms of suitability to the field, language structure, suitability to the student level and clear comprehensibility after obtaining the expert opinions and four items were omitted from the achievement test.

2.3. Application and data analysis

The achievement test consisting of 31 items was composed of multiple-choice questions and applied to 111 students studying in 6th, 7th, and 8th grades in Elazığ Bahçeşehir College. The achievement test was prepared with 4 options. After the application of the test, scoring was made by giving "1" point to the correct answers and "0" point for the wrong or blank answers.

Table 1 presents the professional status and gender of the experts who were consulted.

Table 1. Characteristics of the experts who were consulted

Number	Professional Status	Gender
1	Asst. Prof.	Male
2	Res. Asst.	Male
3	Science Teacher	Male
4	Science Teacher	Male

3. RESULTS

This section contains the results found in the development phases of the achievement test individual difficulty levels of items and discrimination indexes were calculated for the validity analysis of the achievement test for the science lesson within Wedo 2.0 robotic set. Table 2 shows the "P" and "D" values in the analysis results of the test prepared with 31 items.

Table 2. Difficulty degrees (p) and discrimination indexes (d) of the items in achievement test

Question No	Groups	Total Score	P Value	D Value
Q.1	Supergroup	39	0.729729	-0.054545
	Subgroup	42		
Q.2	Supergroup	54	0.918918	0.109090
	Subgroup	48		
Q.3	Supergroup	53	0.837837	0.236363
	Subgroup	40		
Q.4	Supergroup	53	0.855855	0.2
	Subgroup	42		
Q.5	Supergroup	54	0.873873	0.2
	Subgroup	43		
Q.6	Supergroup	49	0.711711	0.345454
	Subgroup	30		
Q.7	Supergroup	51	0.810810	0.218181
	Subgroup	39		
Q.8	Supergroup	50	0.792792	0.218181
	Subgroup	38		
Q.9	Supergroup	42	0.531531	0.454545
	Subgroup	17		
Q.10	Supergroup	41	0.531531	0.418181
	Subgroup	18		
Q.11	Supergroup	46	0.621621	0.418
	Subgroup	23		
Q.12	Supergroup	31	0.495495	0.1272
	Subgroup	24		
Q.13	Supergroup	12	0.171171	0.0909
	Subgroup	7		
Q.14	Supergroup	49	0.711711	0.345454
	Subgroup	30		
Q.15	Supergroup	52	0.756756	0.363636
	Subgroup	32		
Q.16	Supergroup	52	0.765765	0.345454
	Subgroup	33		

Q.17	Supergroup	33	0.450450	0.309090
	Subgroup	16		
Q.18	Supergroup	20	0.288288	0.145454
	Subgroup	12		
Q.19	Supergroup	48	0.756756	0.218181
	Subgroup	36		
Q.20	Supergroup	27	0.468468	0.03636
	Subgroup	25		
Q.21	Supergroup	37	0.504504	0.327272
	Subgroup	19		
Q.22	Supergroup	52	0.648648	0.581818
	Subgroup	20		
Q.23	Supergroup	48	0.648648	0.436363
	Subgroup	24		
Q.24	Supergroup	55	0.747747	0.490909
	Subgroup	28		
Q.25	Supergroup	53	0.720720	0.472727
	Subgroup	27		
Q.26	Supergroup	53	0.792792	0.327272
	Subgroup	35		
Q.27	Supergroup	48	0.693693	0.345454
	Subgroup	29		
Q.28	Supergroup	55	0.837837	0.309090
	Subgroup	38		
Q.29	Supergroup	43	0.558558	0.436363
	Subgroup	19		
Q.30	Supergroup	41	0.486486	0.5090909
	Subgroup	13		
Q.31	Supergroup	54	0.909909	0.127272
	Subgroup	47		

When Table 2 was examined, difficulty levels and discriminative values of the items can be seen. One of the feasibility conditions for achievement test is to separate students who find right and wrong options. When the item difficulty levels (P) were examined, it was seen that items with moderate level (values close to 0.5) were in majority and the achievement test was composed of all kinds of questions from easy to difficult. It was concluded that items whose difficulty level is close to and distant from the value of 0.5 namely easy and difficult items should be added.

The item discrimination index (D) in the achievement test was also analyzed in Table 2. When the literature is examined, it is stated that items are successful in discriminating the students who gave correct and incorrect answers if item discrimination (D) index meets the requirement of $D \geq 0.25$ (Çepni et al., 2008).

There are many criteria in the literature about discrimination index. These criteria are that the item discrimination indexes in the test are between +1 and -1. If the discrimination index value is $0.19 \geq D$, then it is stated that the item is very weak and needs to be omitted from the test. It is stated that if item discrimination value is $0.20 \geq D \geq 0.29$, then the item can be used in necessary conditions; if the value is $0.30 \geq D \geq 0.39$, then the item is quite good but it needs to be fixed and improved, if it meets the condition of $D \geq 0.40$, then the item is very good and it can be used directly. The increase in the validity of the test depends on the high value of the item discrimination (Turgut, 1992; Tekin, 2000).

Table 3. Distribution of the Questions in the Test According to their Discrimination Index Values

D Value		Total
0.40 and higher	Very good	9
0.30-0.39	Good, can be improved	9
0.20-0.29	Should be revised in general	6
0.19 and lower	Weak, should be revised or not included in the test	7

When Table 2 and Table 3 were examined, it was determined that there were more than one kinds of "D" values and items 1, 2, 12, 13, 18, 20, and 31 did not meet the condition of $D \geq 0.25$ according to the literature. Unsuitable items were excluded from the test. Items 6, 9, 10, 11, 14, 15, 16, 17, 21, 22, 23, 24, 25, 26, 27, 28, 29 and 30 that were appropriate to gains and met the condition of "D" values higher than 0.25 were ensured to be definitely used in the test. As a result of the analyses, a total of 24 items with appropriate item difficulty levels and item discrimination indexes were numbered between 1-24 and the test was finalized. The achievement test, structured in 24 items, was made a highly validated test with expert opinions and item analyses.

The difficulty levels of the items in the test varied between .17 - .91. The difficulty levels in this range indicate that there are easy and difficult items in the test. The average difficulty value of the items in the test was determined as .66. When considering that the mean test difficulty value is .50 according to the literature, it can be asserted that the achievement test is a test with "the desired average difficulty degree" in moderate difficulty level (Tekin, 2000). For the reliability of the test, KR-20 analysis was conducted and the reliability coefficient was found as .88. When all the results were examined, it can be asserted that the achievement test was reliable. Table 4 shows the statistical procedures related to the achievement test and the obtained data.

Table 4. Achievement Test Analysis Results

Number (N)	Arithmetic Mean ($\bar{X}$)	Standard Deviation (St)	Mean Difficulty (P)	Reliability KR-20
111	16.65	5.67	0.66	.889

As a result of the data obtained for the overall achievement test, it can be asserted that the test is a valid and reliable test.

4. CONCLUSION AND DISCUSSION

Within the scope of the studies, an achievement test consisting of multiple choice questions with high level of validity and reliability was developed. Test development stages were followed within this purpose. These stages were;

- ✓ Determining the purpose of the use of test scores,
- ✓ Preparing a specification table for the content validity
- ✓ Referring to expert opinion
- ✓ Establishing the first item pool of the test
- ✓ Forming the pilot form of the test, and conducting the pilot test in suitable schools,
- ✓ Determining item and test statistics and removing any unsuitable items,
- ✓ Determining how to score the test.

When the literature is examined, it is determined that there are many studies having these stages and being prepared in the fields of chemistry, physics, basic information technologies and biology (Akgün and Gülmez Güngörmez, 2017; Bingöl and Halisdemir, 2017; Demir, Kızılay and Bektaş, 2016; Açıkgöz and Karslı, 2015; Akbulut and Çepni, 2011; Gönen and et al., 2011; Tekbıyık and Akdeniz, 2010).

This achievement test is a scale having applicable qualities including questions appropriate to the science gains within WeDo 2.0 robotic set. This achievement test is a valid material with ensured validity and reliability in terms of measuring and evaluating the deficiencies of the students on the basis of the target behaviors based on the subject. Today, interdisciplinary studies are becoming more and more important and the need for evaluation increases. When the literature is examined, it is seen that the science studies supported by robotic education are not sufficient and there are no measurement tools. This achievement test, which is created with the aim of measuring the

achievements of secondary school students in the science studies supported by robotics education, is important in terms of eliminating the deficiencies in the field.

It is very important to measure target behaviors, and to develop and use scales for behaviors. Multiple choice questions, which are mostly preferred to measure the success of the students, are one of the leading measurement tools since it enables to measure all gains in a short time and get results (Kempa, 1986; Ogan Bekiroğlu, 2004). With the multiple-choice questions, mislearning of the students and their lack of knowledge are determined. It is the most preferred measurement tool in terms of measuring all required information covering the target gains compared to the other measurement tools. Achievement tests composed of multiple-choice questions are formed and tests whose validity and reliability are provided with the analyses are prepared (Nazlı and Başer, 2008).

The discrimination indexes of the items in the final form of the test were 0.40 and above. The item discrimination index value of 0.40 and above is defined as "Very good item" (Büyüköztürk, 2009). When the literature was examined, it was determined that the item discrimination index of the prepared tests was 0.40 and above. For example, the mean discrimination index of the test developed in the study by Demir, Kızılay and Bektaş (2016) was found as 0.485. Similarly, the mean discrimination index of the achievement test developed in the study by Açıkgöz and Karslı (2015) was calculated as 0.44. These results had similarities with the results of the study. Similarly, the difficulty levels of the items in the final form of the test had the values between .17 and .91. The mean difficulty index of the test was 0.66. Similar results have been obtained from other studies conducted in this field. For example, in their study, Akgün and Gülmez Güngörmez (2017) developed a Scientific Reasoning Test to reveal the reasoning levels of the students for the Science Lesson. The mean difficulty was determined as 0.589 as a result of the test analysis. Similarly, the mean difficulty of the test developed by Bingöl and Halisdemir (2017) was calculated as 0.59. For this reason, it can be stated that the test developed in the study is suitable for the level of the students in terms of the item difficulty. When the studies in the literature are examined, KR-20 and KR-21 methods are seen to be used when calculating the reliability coefficient of the developed achievement tests (Demir, Kızılay and Bektaş, 2016; Özüdoğru and Adıgüzel, 2015; Adıgüzel and Özüdoğru, 2013; Varış and Cesur, 2012). As a result of the internal consistency analysis of the test, the KR-20 coefficient was calculated as .889. KR-20 reliability coefficient value close to 1.00 refers to high reliability of the test and its value close to 0.00 refers to low reliability for the test. In other words, high reliability of the test shows that the error in the test scores is low; whereas, low reliability of the test indicates that the error in the test scores is high (Özçelik, 2010). The final form of the test consists of 24 questions. As a result of the conducted item and internal consistency analysis, it was seen that the test had questions at various difficulty levels and had high reliability and validity values. In the phase of scoring of the test, it was decided to score each correct answer with 1 and each wrong or blank answer with 0.

The achievement test having the content appropriate to the science lesson achievement of Wedo 2.0 robotic set can be an alternative for the researchers as a valid and reliable measurement tool to be used in the future interdisciplinary studies for the science lessons supported with Wedo 2.0 robotic set.

5. RECOMMENDATIONS

- ✓ With this study, an achievement test was developed to measure the target gains of Wedo 2.0 robotic set for the science lesson contents and tests can be developed to measure different interdisciplinary gains with other studies.
- ✓ Tests can be prepared to measure the effects of robot sets such as EV3 and NXT on the science and mathematics lessons.
- ✓ This test, which was prepared for secondary school level, can be created for students at primary school level with different gains.

- ✓ The prepared achievement test is useful due to the information it provides to developers. However, it also has some limitations. Therefore, it is possible to reach various results with different methods and types of questions including different information.
- ✓ The number of people in the application group can be increased to improve the test and more powerful results can be obtained for the analysis.
- ✓ When determining the sample group, students at various levels can be studied.
- ✓ This developed achievement test can be used in similar studies.
- ✓ By focusing on interdisciplinary studies supported with robotic education in education programs, learners should be made aware of such interdisciplinary issues.
- ✓ Various activities can be conducted with teachers and students in order to raise sufficient awareness in the fields of interdisciplinary education programs.

Acknowledgment

This study was supported by the Scientific Research Projects Coordicantion Unit of Fırat University (Project No. EF.18.01).

REFERENCES

Açıkgöz, M. ve Karslı, F. (2015). Validity and Reliability Analysis of Achievement Test Developed on Work and Energy Using Alternative Measurement and Evaluation Techniques, *Journal of Amasya University Faculty of Education*, 4(1), 1-25.

Adıgüzel, O. C. ve Özüdoğru, F. (2013). Developing an Academic Achievement Test for Common Compulsory Foreign Language I Course in Universities, *Trakya University Journal of Education Faculty, Volume 3,* Number *2, 1-11.*

Akgün, A. ve Gülmez Güngörmez, H. (2017). Development of Scientific Reasoning Skills Test for Secondary School Students, *2. International Symposium On Social Sciences,* (18-20 May 2017), Antalya/Alanya.

Bingöl, A. ve Halisdemir, N. (2017). Academic Achievement Test Development Study for Basic Information Technologies Course of University Students, *The Journal of Academic Social Science Studies (JASSS),* Number: 54 , p. 541-554.

Büyüköztürk, Ş. (2009). *Data analysis manual for social sciences: Statistics, research design, SPSS applications and interpretation (10th edition).* Pegem A Publications, Ankara.

Costa, M. F. & Fernandes, J. (2005) Robots at School. The Eurobotice project. *Proceedings of Hsci*2005

Çepni, S. (2005). *Science and Technology Teaching*. Pegem-A Publications: Ankara.

Çepni, S., Bayrakçeken, S., Yılmaz, A., Yücel, C., Semerci, Ç., Köse, E., Sezgin, F., Demircioğlu, G. ve Gündoğdu, K. (2008). *Quantification and Consideration.* Ankara: Pegem Publications.

Çilenti, K. (1985). Science education technology: teaching, program and test development in science.

Demir, N., Kızılay, E. ve Bektaş, O. (2016). Developing Achievement Test on 7th Grade Solutions: Validity and Reliability Study, *Necatibey Faculty of Education Journal of Electronic Science and Mathematics Education (EFMED),* Volume 10, Number 1, June 2016, pages 209-237.

Goldworthy, A. (2000). Teaching Students How to İnvestigate, *Paper Presented at the Annual Meeting of Science Conference*. Nicosia. Cyprus.

Kempa, R. (1986). *Assessment in Science.* Cambridge University Press, Cambridge, London.

MEB. (2006). Primary Science and Technology Course (6th, 7th and 8th Grades) Curriculum, *Head of Board of Education and Training*, Ankara.

Narlı, S., & Baser, N. E. (2008). A Quantitative Research on Developing a Success Test on "Cluster, Correlation, Function" and Describing the Readiness of the First Year Students of the Department of Mathematics in the Department of Mathematics.

Ogan Bekiroğlu, F. (2004). *How Successful ?, Classical and Alternative Measurement and Evaluation Methods and Applications in Physics (1st edition).* Ankara: Nobel Publication Distribution.

Özçelik, D. A. (2010). *Quantification and consideration.* Ankara: Pegem Academy.

Tekin, H. (1992). *Measurement and Evaluation in Education*. Ankara: Yargı.

Varış, Y. A. ve Cesur, D. (2012). Developing Achievement Test for Secondary Music Course, *Turkish Studies - International Periodical For The Languages, Literature And History Of Turkish or Turkic Volume 7/4, Fall 2012, P. 3189-3198*