Development of An Achievement Test For 5th Grade Foreign Language Classes[1]

Derya Yüreğilli GÖĞSU [2*]   Yücel GELİŞLİ [3]

2 Phd.,Gazi Faculty of Education, Department of Educational Sciences, Gazi Üniversitesi Rektörlüğü
Tarihi Bina 06500 Teknikokullar/ Turkey
3 Phd.,Gazi Faculty of Education, Department of Educational Sciences, Gazi Üniversitesi Rektörlüğü
Tarihi Bina 06500 Teknikokullar/ Turkey
* E-mail of the corresponding author: deryagoksu06@gmail.com

Abstract

The aim of the current study is to develop an achievement test of multiple-choice questions that assesses reading skills of Movies and Party Time Units of 5th curriculum. The achievement test was analyzed to form a final test of 25 questions and applied to 234 6th grade students in the 4 different schools of 4 different regions in Ankara during fall semester of 2016-2017 academic year as the 1st pilot trial. Following the 1st pilot trial's item discrimination analysis and item difficulty index analysis, 12 items of the test is revised. After revising 12 items, 2nd pilot trial was conducted to different groups of 135 people of 6th grade students in the same regions of Ankara, After the 2nd Pilot trial, item discrimination and item difficulty analysis results were calculated. Following the analysis, 6 items were exracted from the test . For the internal reliability of the test, KR-20 value was calculated as 0, 924. The mean item difficulty (p) of the test was determined as 0,533 and mean item discrimination index (rjx) was determined as 0,521.

Keywords: *5th grade curriculum, achievement test, item discrimination index, item difficulty index*

1. Introduction

In the scope of an education , the concept of a curriculum has a significant role. The curriculums are the programs that are established according to certain information categories and planned in parallel to certain knowledge and skill education programs in order to provide target audience. (Küçükahmet, 2008) Before starting to apply a curriculum education some exams are performed in order to determine students' preliminary behaviors, to assess curriculum's functionality and the areas which students have trouble, to determine students' learning level and for other similar reasons. Besides, there are some tests applied to assess students' levels when they finish the course. In order to measure whether the students have acquired the course acquisition or not regarding the teaching and learning process and output level of education, monitoring tests, achievement tests or attitude scales and surveys are used. For the Turkish Education System, generally achievement tests are applied to students for evaluating. To evaluate a student's achievement, the results from the achievement tests applied to the student have been taken into consideration. The attitudes desired to ensure the students and determine the acquisition levels are measured by the achievement tests before and after a course-basis. For this reason, it is important for a teacher to gain some technical knowledge about the preparation and development of the achievement test. (Tan, 2008) Because the students' academic achievement level is measured by intellectual activities such as exact remembering, comprehension of the reading and problem solving. Thanks to these technical methods to measure the academic achievement, it is assessed that how the students reflect their learning in courses or in the environment outside the courses. (Baykul, 2000) This situation is a proven fact that students have gained to the trainers about their academic achievement. According to Yaman (2003), the academic achievement is defined as "the whole of knowledge and skills of the students regarding the subjects which are worked on in a certain process. Multiple choice tests, gap filling, pairing, determining true-false or long response classical questions are the most preferred measurement techniques." In addition to this, questionnaires, oral exams, true false tests, multiple choice tests, pairing tests, filling the blanks tests, scales, short response tests, written examinations and open ended questions have been used in order to measure students' achievement and evaluate their acquisition.(Kempa at al.,1986). The reliability and validity calculations are performed after the

course practices that proves the reliability and the validity of the tests. In this context, it can be proved that the test is strong. In the academic achievement tests, the most important subject of is the reliability and the validity of a test. (Büyüköztürk, 2014).

In the 21st century the most demanded language to learn commonly is English. (Fang at al., 2012). In addition to the fact, being accepted as the language of science and the high majority of countries whose native language is English make this language as the most preferred language in foreign language education. (Canbulat & İşgören, 2005). In the scope of foreign language education, in order to evaluate statistically how the language course objectives provided to students can be calculated by the application of valid and reliable tests that will help educators for the right evaluation process. For assessing the students' achievement of foreign language acquisition, rather than using general grammar tests, the tests which measure four main language skills and the self- evaluation of students in the process of teaching-learning is getting important (Demirel, 2005). In order to determine quality and quantity of the inputs acquired by the students, the difficulties experienced in learning, and to measure achievement, the assessment and evaluation process is required. By this way, the defects of the learning process are found and after the determination of achievementes and failures of learners, relevant improvement studies can be applied. In this context measurement tools in foreign language teaching takes a significant place. Because language education is multi-dimensional. Different areas such as skills, language structures, vocabulary knowledge, discourse knowledge are approached. In order to achieve this prospect, the prepared or preferred tool should be valid and reliable. (Snow, 1997) Besides, the tests which are proven to be valid and reliable should be applied to students separately for each skill and then, student should be evaluated. Because there are separate acquisitions for each skill in the foreign language course curriculum. For this reason, student evaluations prepared for each skill should be applied as a separate test. Consequently, evaluation in the foreign language teaching is executed to check what students can do and cannot do with the target language. (Abbott&Greenwood, 1985)

THE PURPOSE OF THE STUDY

The purpose of the current study is to develop a valid and reliable achievement test to measure 5th grade students' level of reading skills regarding the "Movies" and "Parties" Units. This research is important to inform the teachers about how an achievement test can be developed and to what extent they have reached the objectives of their course.

METHOD

Research Design

The current study is a descriptive study. The descriptive study includes the works which were performed in a sample of a universe or a universe which consists of many elements in order to reach a general conclusion about the universe. (Karasar, 1994) In this study, an achievement test that measures two units' objectives was developed for 5th grade students which is also in parallel with the achievements of "Movies" and "Party Time" units within the scope of English Language Teaching Curriculum provided by the Discipline of Ministry of National Education.

Participants

The research is conducted in fall semester of 2016-2017 education year in 4 different state schools of Ankara. Considering that 5th grades have not learned those topics, two pilot trials were applied to the 6th grade students. Four different districts were selected in Ankara for the pilot practices. The selection of schools within the districts was made through convenience sampling.

Data Collection Tools and Data Analysis

To develop a valid and reliable achievement test, firstly a planned process should be followed. According to Turgut and Baykul (2010), the stages of achievement test development are stated as the determination of the goal of the test, the determination of features which will be measured, the writing of the items, reduction of items, the preparation and execution of the trial form, scoring of the execution results, item analysis and finalizing the test by item selection and the statistics. In the current study, these steps are carefully followed and every stage is carried out respectively. In order to develop the data gathering tool, firstly a goal is set regarding why this test

should be used. In this study, the aim of developing an achievement test is to see the objectives for these 2 units of 5th grade in a scientific and solid way. The features which will be measured in the developed achievement test are actually the behaviors which are included in the 5th grade English course's "Movies" and "Party Time" units and stated as the desired reading skills which should be provided to the students within the scope of Ministry of National Education teaching programs. A Statement table has been prepared to assess that these features are reflected equally to the item distributions in general. In the Signal Table, the vertical column includes the subjects of two units, and the rows contain the sections about which items measure which subjects. After the consolidation of the table was created, the original questions are started to be produced after the content was examined for the writing of items in parallel to achievements stated in the table. In addition to the 25 questions which will be included in the final test, after the item analyzes, a pool of 55-items question was created for alternative. Item distribution was applied for each subject with a minimum of 3 and a maximum of 5 questions in the signal table. After items written in adequate amount, for their reduction process, they were presented to English teachers who had completed at least the 10th year in their office in order to check those items validity. Besides, 2 assessment and evaluation specialists and 3 program development specialists were asked about their opinions about the final test and necessary adjustments have been done. After the last adjustments, questions were created and a trial form was prepared. The test form applied as the 1st pilot application in the study universe and then the results were scored. After the analyzes in the TAP program, item difficulty index and item discriminants were determined and after adjusting the inappropriate items, the second pilot trial is implemented. After the second pilot test, the items in the final test is determined by categorizing the items according to their difficulty level.

FINDINGS

In this section, findings which are deducted from validity and reliability studies of the tests are mentioned. Item analysis were performed by using the TAP program to ensure the validity of the test item structure. The KR-20 value was checked for internal reliability of the test and it was concluded that the test has an adequate reliability (0.924). The average item difficulty (p) was determined as 0.533 and the test's average item discrimination index (rjx) was determined as 0.521. The statistical analyzes proved that the achievement test for 5th grade is a valid and reliable test. Item discrimination indices and difficulty ratings are indicated in Table 1. According to item analysis calculations, the items that have discrimination index between 0,20 and 0,30 are usually available for testing. Items with an item discrimination between 0,30 and 0,40 are seen as good; and the items which have discrimination index above 0.40 are accounted as perfect. Items which have discrimination index below 0,20 should be used after they are developed. (Özçelik, 2013). As a consequence of the 1st pilot trial, the items 1., 3., 5., 6., 12., 18., 25., 26., 29., 40., 43 and 48 have a discrimination index below 0.30. Because of the fact they are readjusted and were reapplied in the 2nd pilot trial. After these items were adjusted, the second pilot trial was applied to develop the final achievement test. After the second pilot trial, the items with the number of 14, 23, 16, 29, 30, and 33 that have item index below 0,20 were extracted from test content. Table.1 below shows the number of correct answers given by the students in the upper and lower group, item difficulty index and item discriminations

In Table 1, items that have a discrimination index above 0.30 are indicated. In determining the item difficulty level, the below criteria were taken into consideration and the item selection was performed in relation to the signal table. Considering the item difficulty coefficient, the items which should be included in the final test should be selected as 2% very easy, 13% easy, 70% moderate, 13% difficult and 2% very difficult (Turgut&Baykul, 2012). The item difficulty index (p) is the ratio of true answers to the all subjects of the test. The item difficulty index is evaluated from 0 to 1. Ranging of difficulty index closer to 1 indicates that the item is easy, getting closer to the 0 indicates that the item is difficult (Turgut&Baykul, 2012). Items are classified according to their difficulty indices: between 0.20 and 0.40 is stated as difficult; between 0,41 and 0,60 is stated as moderate and between 0,61 and 0,80 is stated as easy. (Kutlu, 2004) Based on the information that is received by Turgut, Baykul and Kutlu, for the final achievement test development, 15% easy items, 70% moderate difficulty items and finally 15% difficult items are used in this study. The results gathered from this case is indicated in Table 2.

Table 1.
Item analysis according to the number of true answers given by the students in the upper and lower group

item no	p	rjx	dü	da	item no	p	rjx	dü	da
1	0,75	0,51	34 (0,94)	17 (0,44)	28	0,79	0,54	36 (1,00)	18 (0,46)
2	0,66	0,56	35 (0,97)	16 (0,41)	31	0,59	0,61	34 (0,94)	13 (0,33)
3	0,68	0,64	36 (1,00)	14 (0,36)	32	0,58	0,61	33 (0,92)	12 (0,31)
4	0,48	0,39	27 (0,75)	14 (0,36)	34	0,61	0,66	34 (0,94)	11 (0,28)
5	0,61	0,56	32 (0,89)	13 (0,33)	35	0,64	0,59	34 (0,94)	14 (0,36)
6	0,44	0,76	30 (0,83)	3 (0,08)	36	0,59	0,76	34 (0,94)	7 (0,18)
7	0,51	0,45	29 (0,81)	14 (0,36)	37	0,73	0,64	36 (1,00)	14 (0,36)
8	0,50	0,66	32 (0,89)	9 (0,23)	38	0,34	0,41	23 (0,64)	9 (0,23)
9	0,56	0,74	33 (0,92)	7 (0,18)	39	0,47	0,57	29 (0,81)	9 (0,23)
10	0,57	0,66	35 (0,97)	12 (0,31)	40	0,45	0,54	27 (0,75)	8 (0,21)
11	0,52	0,69	33 (0,92)	9 (0,23)	41	0,60	0,55	29 (0,81)	10 (0,26)
12	0,65	0,40	31 (0,86)	18 (0,46)	42	0,62	0,66	35 (0,97)	12 (0,31)
13	0,67	0,64	35 (0,97)	13 (0,33)	43	0,36	0,33	21 (0,58)	10 (0,26)
15	0,70	0,56	33 (0,92)	14 (0,36)	44	0,53	0,63	31 (0,86)	9 (0,23)
16	0,83	0,38	35 (0,97)	23 (0,59)	45	0,56	0,50	28 (0,78)	11 (0,28)
17	0,59	0,55	29 (0,81)	10 (0,26)	46	0,57	0,60	31 (0,86)	10 (0,26)
18	0,32	0,35	20 (0,56)	8 (0,21)	47	0,48	0,47	27 (0,75)	11 (0,28)
19	0,43	0,33	24 (0,67)	13 (0,33)	48	0,42	0,41	23 (0,64)	9 (0,23)
20	0,69	0,56	34 (0,94)	15 (0,38)	49	0,41	0,49	27 (0,75)	10 (0,26)
21	0,46	0,44	27 (0,75)	12 (0,31)	50	0,64	0,58	33 (0,92)	13 (0,33)
22	0,44	0,41	22 (0,61)	8 (0,21)	51	0,58	0,61	33 (0,92)	12 (0,31)
24	0,56	0,66	32 (0,89)	9 (0,23)	52	0,48	0,68	31 (0,86)	7 (0,18)
25	0,47	0,34	26 (0,72)	15 (0,38)	53	0,65	0,69	36 (1,00)	12 (0,31)
26	0,56	0,79	32 (0,89)	4 (0,10)	54	0,48	0,71	33 (0,92)	8 (0,21)
27	0,64	0,79	36 (1,00)	8 (0,21)	55	0,33	0,49	23 (0,64)	6 (0,15)

As indicated in Table 1

dü: The number of correct responses in the upper group students,
da: The correct number of students in the below group,
p: Item difficulty,
rjx: Item discrimination.

Table 2.
The item diffuculty index of the items

Item diffuculty	Item number	Item No	Conclusion
$0,61 \leq p \leq 0,80$	16	1,2,3,5,12,13,15,20,27, 28,34,35,37,42,50,53	Easy Items
$0,41 \leq p \leq 0,60$	29	4,6,7,8,9,10,11,17,19,21,22,24, 25 26,31,32,36, 39,40,41, 44,45,46,47,48,49,51,52,54	Moderate Items
$0,20 \leq p \leq 0,40$	4	18,38,43,55	Dıffucult Items

The item numbers' item difficulty indices given in Table 2 are in the range of $0,81 \leq p \leq 0,20$. The items listed in Table 1 do not include items 14, 16, 23, 29, 30 and 33. These items were removed because they have item discrimination index was below 0.20 ($p \leq 0.20$) and their item difficulty were not in the range of $0.81 \leq p \leq 0.20$. Hence, the number of items which can be used in the test decreased to 49 from 55. The items which can be used are determined after the distribution of topics and objectives of the Units. The final test items are shown in the Statement Table that is indicated in Table 3. The item numbers of the test which is decreased to 49 items were stated separately in the signal table. The item numbers corresponding to subjects in the statement table are distributed regarding the categories easy, medium and difficult. After the analyzes and index calculations performed, the final test consisting 25 items was developed for the main trial.

Table 3.
The Objective Statement Table

	OBJECTIVES	
UNIT 6-MOVIES	Students will be able to comprehend phrases and simple sentences in short texts, on posters and advertisements related to movies.	Total
Describing characters/people	4- 9	2
Describing what people do	31- 34	2
Expressing likes and dislikes	6-54	2
Making simple inquiries	18	1
Stating personal opinions	11-10	2
Telling the time, days and dates	17	1
Total		10
UNIT 7-PARTY TIME	Students will be able to understand short texts with visual aids such as cartoons, posters and birthday cards.	Total
Asking for permission	38-52	2
Expressing and responding to thanks	39-46	2
Expressing basic needs	42-55	2
Expressing likes and dislikes	32	1
Expressing obligation	27-35	2
Greeting and meeting people	43-51	2
Making simple inquiries	44-53	2
Telling the time, days and dates	22-26	2
Total		15

5. Conclusion

In this study, it was aimed to develop an achievement test that consists of multiple choice questions which are valid and reliable in accordance with the acquisitions of the reading skills provided in the "Movies" and "Party Time" units of the 5th grade foreign language teaching program. Multiple choice questions inform us about the topics that students are having trouble It offers possibility to measure and evaluate all information regarding any subject or unit. It is the most preferred measurement tool to evaluate student achievement compared to other measurement evaluation tools in teaching profession. The achievement tests that have been worked on are in the process of standardization in order to obtain tests that have high degree of validity and reliability. (Narlı&Başer, 2008).

Many measurement and evaluation tools can be developed to determine to what extent the targeted acquisitions have been realized in education. Validity and reliability determinations are crucial to develop achievement tests with the purpose of using in academic achievement evaluation. In order to measure target realization of desired reading skills in foreign language education, the validity and reliability of the "English Achievement Test" was examined; and as a result, it is concluded that the developed test is a valid and reliable measurement tool. The final version of English Achievement Test is provided in the Appendix. The final version of the test in attachment is a document which assesses the target reading skills regarding the "Movies" and "Parties" units in 5th Grade English Language education program and it is in a capacity that enables English teachers to use it. In addition to this, the achievement test developed within the study provides an example for educators to develop their own assessment tools.

It is a crucial point to decide which methods and techniques that will be used in order to provide necessary acquisitions to the target group education and teaching process. But also the analyzing the extent which targets are achieved holds importance as much as the process itself. Measuring and evaluation techniques and tools should be selected correctly and carefully by educators and any kind of mistakes should not be ignored. The assessment tools which will be applied at the end of education-teaching program should be suitable the level of the target group, they should be controlled through the validity and reliability studies in accordance with the achievements, and relevant development studies for these assessment and evaluation tools are performed and enhanced.

References

Abbott, G. & J. Greenwood, (1985). *The Teaching of English as an International Language*, Collins: Glasgow and London.

Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme, Klasik Test Teorisi ve Uygulaması.* Ankara: ÖSYM Yayınları.

Büyüköztürk, Ş. (2014). Sosyal bilimler için veri analizi el kitabı. Ankara: PEGEM Akademi.

Canbulat, M. & İşgören, O.Ç. (2005): "Yabancı Dil Öğretimine Başlamada Kullanılan En Uygun Yaşın Ne Olduğuna İlişkin Dilbilimsel Yaklaşımlar ve Öğretmen Görüşleri", Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, Ankara: Detay Kitap ve Yayıncılık

Demirel, Ö. (2005): "Avrupa Konseyi Dil Projesi ve Türkiye Uygulaması", Milli Eğitim Üç Aylık Sosyal Bilimler Dergisi, Ulaşım tarihi: 20.07.2006 (http://yayim.meb.gov.tr/dergiler/167/orta3-demirel.htm)

Fang, D.U. (2012). How to develop students' communicative competence in spoken English teaching. *Journal of Hunan Industry Polytechnic*, 4, 040.

Jacobsen, C. (2015). The impact of listening strategy instruction on the learning of English and an additional foreign language. *Malaysian Journal of ELT Research*, *11*(1), 17.

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

Ogan Bekiroğlu, F. (2004). *Ne kadar Başarılı?, Klasik ve Alternatif Ölçme- Değerlendirme Yöntemleri ve Fizikte Uygulamalar* (1. baskı). Ankara: Nobel Yayın Dağıtım.

Küçükahmet, L. (2008). Öğretimde Planlama ve Değerlendirme. Ankara. Nobel Yayınları.

Liao, F.Y. (2012). *ESL students' writing experiences and perceived poetry writing ability* (Yayımlanmamış Yüksek Lisans Tezi). Indiana University of Pennsylvania, Pensilvanya.

Narlı, S. & Başer, N. (2008). "Küme, Bağıntı, Fonksiyon" Konularında Bir Başarı Testi Geliştirme ve Bu başarı testi İle Üniversite Matematik Bölümü 1. Sınıf Öğrencileirnin Bu Konulardaki Hazırbulunuşluklarını Betimleme Üzerine Nicel Bir Araştırma. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, 24, 147-158.

Nepomuceno Reyes, A. (2014). *Effective strategies and techniques for teaching English at a nursery school* (Yayımlanmamış Yüksek Lisans Tezi). Universidad Veracruzana, Veracruz.

Snow, M. A., D. M. Brington, (1997). The Content–Based Classroom: Perspectives on Integrating Language and Content, Longman.

Şimşek, A. (2009). *Öğretim Tasarımı (1. baskı),* Ankara: Nobel Yayın Dağıtım.

Tan, Ş. (2008). Eğitimde Ölçme ve Değerlendirme (1. Baskı). Ankara: Pegem Yayınları.

Turgut, M.F. & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme.* Ankara: Pegem Akademi.

Yılmaz, H. (2004). *Eğitimde Ölçme ve Değerlendirme* (7. baskı). Konya: Çizgi Kitabevi Yayınları.

Yaman, S. (2003). *Fen Bilgisi Eğitiminde Probleme Dayalı Öğrenmenin Öğrenme Ürünlerine Etkisi,* Yayınlanmamış Doktora Tezi, Ankara: Gazi Üniversitesi Eğitim Bilimleri Enstitüsü