Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.6, No.24, 2015

Development of Achievement Test: Validity and Reliability Study for Achievement Test on Matter Changing

Filiz Kara* Dilek Çelikler
Department of Elementary Science Education, Ondokuz Mayıs University, Samsun, Turkey

Abstract
For "Matter Changing" unit included in the Secondary School 5th Grade Science Program, it is intended to develop a test conforming the gains described in the program, and that can determine students' achievements. For this purpose, a multiple-choice test of 48 questions is arranged, consisting of 8 questions for each gain included in the training program. The test, of which the content validity is reviewed and ensured by 2 chemistry domain experts and 2 science lecturers, is applied to 354 6th grade students (ages 11-12) in the Black Sea Region of Turkey, in a city centre. Item analysis of the test is carried out and 16 items of which the distinctiveness are below 0,30 are excluded from the test. As a result of item analysis the average difficulty of the questions are estimated to be 0,38 and it is seen that their difficulty level is intermediate. Likewise, the average distinctiveness of the questions are estimated to be 0,38 and it is seen that the distinctiveness strength of the questions are well. After the questions are excluded, Kuder Richardson-20 reliability coefficient is estimated to be 0,763. As a result of the study, an effective and reliable achievement test including 32 questions with intermediate difficulty level and well distinction strength created for "Matter Changing" unit is brought to the science education.
Keywords: Validity, reliability, matter changing, science.

1. Introduction

Exams are the important assessment tools that are utilized at every phase of the education system. Assessment is digitalization of the qualifications, expressing the observed qualifications via numbers and symbols. Evaluation, on the other hand, is a decision making process relating to the assessed qualification, by comparing the results obtained from the assessment process with certain criteria (Özçelik, 1992). The general objectives of the assessment and evaluation processes performed in schools are listed below:

1. To determine the level of forwardness of students to the course,
2. To determine of how much the students possess the behaviours to be taught in the course,
3. To determine how much the students achieved to the gains of the program and how much the learning occurred, at the end of the unit,
4. To notify the students about their deficiencies, at the end of the unit (Özçelik, 1998),
5. To assess the students' skills at the end of the course,
6. To motive the students to different courses and learning fields,
7. To evaluate the validity of an education program or of a method of such program (Kempa, 1986; Yılmaz, 2004).

Often the tests are the assessment tools that are used for determination of the students' gains relating to the cognitive domain within the quantitative researches of education (Sönmez & Alacapınar, 2013). Oral examinations, true-false tests, multiple-choice tests, matching tests, fill-in-the-blank exams, scales, short answer tests, written examinations, open ended questions, two phase testing are used in order to assess and evaluate the achievement of the student at all the stages and in all the fields of the education (Kempa, 1986; Ogan Bekiroğlu, 2004; Yılmaz, 2004; Şimşek, 2009). These all test methods have superior or weak aspects involved compared to each other. According to the researches, multiple-choice tests are the most common method, following the interviews, on revealing the students' knowledge on a specific concept or subject (Kempa, 1986; Ogan Bekiroğlu, 2004).

Multiple-choice tests are the tests those which have only one true answer which is selected from within other obfuscatory answers (Öncü, 1999). Multiple-choice tests are the tests with objective grade which does not tend to differentiate from person to person (Gronlund & Lind, 1990) and are able to be graded in a short time. These tests also allow for a comprehensive evaluation to be made and, with ably written items, for assessing high-level talents (Worthen, Borg & White, 1993). 2-3 choice questions are suitable for the first or second grade students of the elementary school, whereas 3-4 choice questions are suitable for the following grades (Turgut & Baykul, 2010).

The validity is degree of the test's ability to gather the information on the quality that is intended to be assessed (Kaptan, 1998). Reliability on the other hand, is the cohesion between the answers given to the test items. Reliability of a test depends on two main criteria. Which are; cohesion between the answers given in different times and cohesion between the answers given in the same time (Büyüköztürk, 2004).

Whether a test is reliable or not, can be established through various ways. Such as: test-re-test reliability, parallel (equivalent) form reliability, two semi-test reliability, Kuder Richardson-20 (KR-20) and

Cronbach alpha (α) reliability. For tests those of which the item analysis are performed, the reliability coefficient is often determined via KR-20. The KR-20 is used in order to review the internal consistency between the points obtained from the test applied at the same time. If the difficulty levels of the items in the test are close to each other as a result of the item analysis, then KR-21 can be used instead of KR-20 (Büyüköztürk, 2004). Reliability coefficient values can be between 0.00 and 1.00, but cannot be negative (Fraenkel & Wallen, 2009). A test with reliability coefficient 0,70 and above, is usually considered satisfying in terms of reliability (Büyüköztürk, 2004; Fraenkel & Wallen, 2009).

The "Matter Changing" unit for the 5th grade of the secondary school includes the subjects change of matter state (change of state, melting, freezing, boiling, vaporization, condensation, sublimation, hoarfrost concepts), distinctive features of the matter (melting point, freezing point, boiling point concepts), heat and temperature (heat, temperature, heat exchange concepts), and the temperature affects the matters (expansion, shrinkage concepts). In the Science Education Program which is revised in 2013, there are 6 gains for this unit (MEB, 2013). These are the subjects of which the students have misconception at most (Erickson, 1979; Osborne & Cosgrove, 1983; Bar & Travis, 1991; Bar & Galili 1994; Stavy 1990, Taber, 2000, Tytler, 2000; Bayrakci, 2007; Gürdal Kazancıoğlu, 2008).

In the study, it is aimed to develop an achievement test of which the validity and the reliability are ensured, which can be used to determine the achievements and of the 5th grade secondary school students (ages 10-11) during the education process and their forwardness to education for the "Matter Changing" unit, in accordance with gains of the Secondary School 5th grade Science Education Program.

2. Method

A test of 48 questions in total consisting of 8 questions for each gain of the program has been prepared for the "Matter Changing" unit, through various literature review by the researchers. The questions in the test are prepared by taking the age group of the students into consideration, so that each question consisting of 4 options will include one true and 3 obfuscatory answers.

Content validity of the test is determined via the opinions of 2 domain experts and 2 science lecturers.

After the test is applied to students, item analysis of the test is carried out by calculating the difficulty and distinctiveness of the questions of the test, validity and reliability survey is performed, inappropriate questions are excluded, KR-20 reliability coefficient is calculated and the test achieved its final form.

The items are evaluated according to difficulty levels (Baykul, 2000; İşman & Eskicumalı, 2003) provided in Table 1 and distinctiveness criteria (Özçelik, 1997; Tekin, 2000) provided in Table 2.

Table 1. Difficulty Levels of Items (Baykul, 2000; İşman & Eskicumalı, 2003).

Difficulty of Item (p)	Assessment of Item
0.70 - 1.00	too easy
0.50 - 0.69	easy
0.30 - 0.49	intermediate difficulty
0.29 and lower than 0.29	too difficulty

Table 2. Distinctiveness Criteria of Items (Özçelik, 1997; Tekin, 2000).

Distinctiveness of item (r)	Assessment of Item	Usage of Item
higher than 0.40	very well item	not needed to be correct
between 0.30 - 0.40	well item	not needed to be correct
between 0.20 - 0.29	distinctiveness is intermediate	can be used in compulsory situation or needed to be corrected
0.19 and lower than 0.19	too weak item (distinctiveness is weak)	cannot be used or need to be corrected again

354 6th grade students studying in 3 different schools in the city centres of the Black Sea Region of Turkey, are participated in the survey. In the development of the achievement test, 6th grade students are preferred those who have learnt the subject before.

3. Findings

3.1. Findings Regarding the Validity of the Test

The multiple choice test developed, is reviewed by 2 chemistry domain experts and 2 5th grade secondary school science lecturers. As a result of the review by the domain experts and lecturers, it is stated that the content validity of the test has been provided, and is suitable for the purpose and 5th grade student level. Taking the suggestions in the result of the review, the test has been made ready to be implemented by making minor corrections on some questions.

3.2. Findings Regarding the Reliability of the Test

After the test has been applied to 354 students, true answers of each student is coded as 1 and false and void answers as 0, and proceeded on the analysis of the test items.

The scores the students get are calculated and sorted by highest to low. The supergroup is designated by selecting the 27% (354*27/100=96 students) of the top rated students according to the test scores, and the subgroup by selecting 27% (354*27/100=96 students) of the lowest rated students. Item difficulty is determined by means of the $p = (Dü + Da)/2N'$ formula (Turgut, 1997), and item distinctiveness through the $r = (Dü - Da)/N'$ formula (Özcelik, 1997) (N: 27% of all students, Dü: number of the supergroup students those who gave the correct answer to the item, Da: number of the subgroup students those who gave the correct answer to the item).

Item analysis results of the achievement test are provided in Table 3.

Table 3. Item Analysis Results of the Achievement Test

Question Number	Dü	Da	p	r	Expression (according to p)	Expression (according to r)	Assessment
1	42	12	0.28	0.31	too difficulty	well	used
2	69	23	0.48	0.48	intermediate difficulty	very well	used
3	67	23	0.47	0.46	intermediate difficulty	very well	used
4	46	19	0.34	0.28	intermediate difficulty	intermediate	not used*
5	56	25	0.42	0.32	intermediate difficulty	well	used
6	65	37	0.53	0.29	easy	intermediate	not used*
7	66	33	0.52	0.34	easy	well	used
8	58	20	0.41	0.40	intermediate difficulty	well	used
9	44	15	0.31	0.30	intermediate difficulty	well	used
10	43	21	0.33	0.23	intermediate difficulty	intermediate	not used*
11	55	17	0.38	0.40	intermediate difficulty	well	used
12	24	13	0.19	0.12	too difficulty	too weak	not used*
13	65	23	0.46	0.44	intermediate difficulty	very well	used
14	56	27	0.43	0.30	intermediate difficulty	well	used
15	60	21	0.42	0.41	intermediate difficulty	very well	used
16	42	12	0.28	0.31	too difficulty	well	used
17	25	12	0.19	0.14	too difficulty	too weak	not used*
18	39	10	0.26	0.30	too difficulty	well	used
19	48	18	0.34	0.31	intermediate difficulty	well	used
20	58	19	0.40	0.41	intermediate difficulty	very well	used
21	20	15	0.18	0.05	too difficulty	too weak	not used*
22	16	10	0.14	0.06	too difficulty	too weak	not used*
23	51	14	0.34	0.39	intermediate difficulty	well	used
24	30	12	0.22	0.19	too difficulty	too weak	not used*
25	60	34	0.49	0.27	intermediate difficulty	intermediate	not used*
26	64	15	0.41	0.51	intermediate difficulty	very well	used
27	58	13	0.37	0.47	intermediate difficulty	very well	used
28	65	15	0.42	0.52	intermediate difficulty	very well	used
29	52	11	0.33	0.43	intermediate difficulty	very well	used
30	46	15	0.32	0.32	intermediate difficulty	well	used
31	58	25	0.43	0.34	intermediate difficulty	well	used
32	40	17	0.30	0.24	intermediate difficulty	intermediate	not used*
33	51	16	0.35	0.37	intermediate difficulty	well	used
34	53	19	0.38	0.35	intermediate difficulty	well	used
35	37	18	0.29	0.20	too difficulty	intermediate	not used*
36	59	19	0.41	0.42	intermediate difficulty	very well	used
37	44	21	0.34	0.24	intermediate difficulty	intermediate	not used*
38	50	19	0.36	0.32	intermediate difficulty	well	used
39	34	20	0,28	0.15	too difficulty	too weak	not used*
40	54	19	0.38	0.37	intermediate difficulty	well	used
41	29	14	0.22	0.16	too difficulty	too weak	not used*
42	53	21	0.39	0.33	intermediate difficulty	well	used
43	68	30	0.51	0.40	easy	well	used
44	62	24	0.45	0.40	intermediate difficulty	well	used
45	25	12	0.19	0.14	too difficulty	too weak	not used*
46	42	12	0.28	0.31	too difficulty	well	used
47	35	20	0.29	0.16	too difficulty	too weak	not used*
48	46	17	0.33	0.30	intermediate difficulty	well	used

*Dü: number of supergroup students those who gave the correct answer to the item, Da: number of subgroup students those who gave the correct answer to the item, p: difficulty index, r: distinctiveness index. *Ouestions*

that are not used

As a result of the item analysis, the 16 questions of which the distinctiveness are lower than 0.30 are eliminated (questions 4, 6, 10, 12, 17, 21, 22, 24, 25, 32, 35, 37, 39, 41, 45, 47) and the test took its final form so that it include 32 questions in total. Since the difficulty levels of the items used in the test are not close to each other, the reliability of the test is assessed via KR-20. The KR-20 coefficient of the test is calculated via the formula

$$KR_{20} = \frac{K}{K-1}\left[1-\frac{\sum pq}{S_x^2}\right]$$ (K: number of items in the test, S: standard deviation).

Descriptive statistics obtained from the test consisting of 32 questions, after excluding the 16 questions, are given in Table 4.

Table 4. Descriptive Statistics Values of the Achievement Test Data

Definitions	Values
Number of item	32
Number of student	354
Mean	11.42
Standard deviation	5.27
Skewness	0.980
Kurtosis	0.874
Medium item difficulty	0.38
Medium item distinctiveness	0.38
KR-20 reliability coefficient	0.763

Average item difficulty is estimated to be 0,38 as a result of the item analysis, and it is concluded that the difficulty of test items is intermediate. Average item distinctiveness is estimated to be 0,38 and the distinctiveness strength of the test items considered well. KR-20 reliability coefficient of the test is estimated to be 0.763.

Number of the questions relating to the gains for the "Matter Changing" unit, before and after the item analysis, are given in Table 5.

Table 5. Number and Distribution of the Questions Prepared Relating to the "Matter Changing" Unit within the 2013 Science Class Training Program, Before and After the Achievement Test Item Analysis

Subjects	Gains	Number of Gains	Number of Questions Before Analysis	Number of Questions After Analysis
Matter Changing of Matter	Performs experiments regarding to the Matter Changing of matters with the effect of the temperature, makes inferences based on the data obtained. *Specifies that the fluids can be vaporized in all temperatures, and describes the basic difference between the vaporization and the boiling.*	1	8	6
Distinctive Features of the Matter	As a result of the experiments, determines the melting, freezing and boiling points, which are the distinctive features of the matter.	1	8	6
Heat and Temperature	Describes the main differences between heat and temperature.	2	8	4
	Performs experiments regarding to the heat transition as a result of mixing fluids with different temperatures and interprets the results.		8	6
Heat Affects the Matter	Performs experiments on expansion and contraction of the matter under the effect of the heat and discusses the results.	2	8	5
	Distinguishes the relation between the expansion and contraction, through the examples from the daily life.		8	5
	Total	6	48	32

The test before analysis was including 8 questions for each subject from the "Matter Changing" unit. However,

after the item analysis the test took its final form so that it include 32 questions of total, consisting of 6 questions for each “Changing of Matter State" and "Distinctive Features of Matter" subjects, and 10 questions for each "Heat and Temperature" and "Heat Affects Matter" subjects.

4. Conclusions and Suggestions

One of the fundamental elements of a successful chemistry education is a successful assessment process. In order to carry out a successful assessment, a test with validity and reliability are ensured is required to be used. For this reason, it is aimed to develop an achievement test for "Matter Changing" unit.

Regarding to the content validity of the test, it is common in the literature that the opinions of the domain experts and lecturers to be consulted (Peterson & Treagust, 1989; Abraham, Williamson & Westbrook, 1994; Ayas & Demirbaş, 1997; Acar & Yaman, 2011; Hürcan & Önder, 2012). Validity of the test which is prepared within the frame of the survey, is ensured in line with the opinions of 2 chemistry domain experts and 2 science lecturers. As a result of the review by the domain experts and lecturers, it is determined that the content validity of the test has been provided, and is suitable for the purpose and student level.

Özçelik (1997) and Tekin (2000) suggested that the items of which the distinctiveness is 0,19 and below should not be used or be reformulated, whereas that the items of which the distinctiveness is between 0,20-0,29 can be used as is in unavoidable circumstances, or should be corrected. Therefore, 16 questions of which the distinctiveness are below 0,30 are excluded from the test, as a result of the item analysis of the test. The final state of the test consists of 32 questions in total, 4 to 6 for each gain in the program.

After the questions are excluded from the test, it is determined that the average item difficulty of the questions is intermediate (0.38) and the distinctiveness is in a well state (0.38).

It is established that the KR-20 coefficient of the test prepared is sufficient (0.763) for the reliability of a test (Büyüköztürk, 2004; Fraenkel & Wallen, 2009).

As a result of the survey, a valid and reliable, multiple-choice test consisting of 32 questions of which the difficulty and distinctiveness is at demanded level for "Matter Changing" unit has been brought to the science education. The developed test is suggested to be used by the science lecturer in order to determine the forwardness of the 5th grade secondary school students to the "Matter Changing" unit training, their achievement during the training process and their misconception. In addition, the test can be used by researchers those who carry out works regarding to the effect of a certain method to the achievements relating to this unit.

This research which is a part of first author’s doctorate thesis is supported by the Ondokuz Mayıs University, Project Management Office with number of PYO.EGF.1904.13.009.

References

Abraham, M. R., Williamson, V. M. & Westbrook, S. L. (1994). A cross-age study of the understanding five concepts, *Journal of Research in Science Teaching*, 31 (2), 147-165.

Acar, B. & Yaman, M. (2011). Bağlam temelli öğrenmenin öğrencilerin ilgi ve bilgi düzeylerine etkisi, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 40, 01-10.

Ayas, A. & Demirbaş, A. (1997), Turkish secondary students’ sonception of introductory chemistry concepts, *Journal of Chemical Education*, 74 (5), 518-521.

Bar, V. & Gaglili, I. (1994). Stages of children’s views about evaporation, *International Journal of Science Education*, 16, (2), 157-174.

Bar, V. & Travis, A. S. (1991). Children's views concerning phase changes, J*ournal of Research in Science Teaching,* 28 (4), 363-382.

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

Bayrakci, M. (2007). İlköğretim 5. Sınıf Öğrencilerinin “Maddenin Değişimi ve Tanınması” Ünitesindeki Temel Kavramları Anlama Seviyeleri ve Oluşan Kavram Yanılgılarının Tespiti, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 60 s.

Büyüköztürk, S. (2004). *Sosyal Bilimler için Veri Analizi El Kitabı*. Ankara: Pegema Yayıncılık, 4th edition, 195 p.

Erickson, G. L. (1979). Children's conceptions of heat and temperature, *Science Education*, 63 (2), 221-230.

Fraenkel, J. R. & Wallen, N. E. (2009). *How to Design Evaluate Research in Education*. New York: McGraw-Hill Companies, 7th edition.

Gronlund, N. E. & Linn, R. L. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company, 6th edition.

Gürdal Kazancıoğlu, H. (2008). İlköğretim 5. Sınıf Fen ve Teknoloji Dersi, Maddenin Değişimi ve Tanınması Ünitesinde Öğrencilerde Oluşan Kavram Yanılgılarının Tespitinde İki Aşamalı Soruların Kullanılabilirliği Üzerine Bir Araştırma, Celal Bayar Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek

Lisans Tezi, 90 p.
Hürcan, N. & Önder, İ. (2012). İlköğretim 7. sınıf öğrencilerinin fen ve teknoloji dersinde öğrendikleri fen kavramlarını günlük yaşamla ilişkilendirme durumlarının belirlenmesi, X. Ulusal Fen Bilimleri ve Matematik Kongresi, Türkiye: Niğde.
İşman, A. & Eskicumalı, A. (2003). *Eğitimde Planlama ve Değerlendirme*. İstanbul: Değişim Yayınları.
Kaptan, S., 1998. *Bilimsel Araştırma ve İstatistik Teknikleri*. Ankara: Tekışık Ofset Yayınları. 11th edition.
Kempa, R. (1986). *Assessment in Science*. London, Cambridge: London Cambridge University Press.
MEB, (2013). *İlköğretim Kurumları (İlkokullar ve Ortaokullar) Fen Bilimleri Dersi (3, 4, 5, 6, 7 ve 8. Sınıflar) Öğretim Programı*, Ankara: Milli Eğitim Bakanlığı Yayınları, 47 p.
Peterson, R. F. & Treagust, D. F. (1989). Grade-12 students' misconceptions of covalent bonding and structure, *Journal of Chemical Education*, 66 (6), 459-460.
Stavy, R. (1990). Children's conception of changes in the state of matter from liquid (or solid) to gas. *Journal of Researcb in Science Teacbing,* 27 (3), 247-266.
Sönmez, V. & Alacapınar, F. G.. (2013). *Örneklendirilmiş Bilimsel Araştırma Yöntemleri*, Ankara: Anı Yayıncılık, 2nd edition, 384 p.
Şimşek, A. (2009). *Öğretim Tasarımı*, Ankara: Nobel Yayın Dağıtım, 1st edition.
Taber, K. S. (2000). Finding the optimum level of simplification: the case of teaching about heat and temperature, *Physics Education*, 35 (5), 320–325.
Tekin, H. (2000). *Eğitimde Ölçme ve Değerlendirme*, Ankara: Yargı Yayınevi, 14th edition.
Turgut, M. F., 1997. *Eğitimde Ölçme ve Değerlendirme*, Ankara, 10th edition.
Turgut, M. F. & Baykul, Y. (2010). *Eğitimde Ölçme ve Değerlendirme.* Ankara: Pegem Yayıncılık, 425 p.
Tytler, R. (2000). A Comparison of Year 1 and Year 6 Students' Conceptions of Evaporation and Condensation: Dimensions of Conceptual Progression, *International Journal of Science Education*, 22 (5), 447-467.
Ogan Bekiroğlu, F. (2004). Ne Kadar Başarılı? Klasik ve Alternatif Ölçme- Değerlendirme Yöntemleri: Fizikte Uygulamalar, Ankara: Nobel Yayın Dağıtım, 1st edition.
Osborne, R. J. & Cosgrove, M. M. (1983). Children's Conceptions of the Changes of State of Water, *Journal of Research in Science Teaching*, 20 (9), 825-838.
Öncü, H. (1999). *Eğitimde Ölçme ve Değerlendirme*, Ankara: Yaysan A. Ş., Geliştirilmiş ve Gözden Geçirilmiş 3rd edition.
Özçelik, D. A. (1992). *Ölçme ve Değerlendirme*, Ankara: ÖSYM Yayınları.
Özçelik, D. A. (1997). *Test Hazırlama Kılavuzu*, Ankara: ÖSYM Yayınları, 3rd edition.
Özçelik, D. A. (1998). *Ölçme ve Değerlendirme*. Ankara: ÖSYM Yayınları.
Worthen, B. R., Borg, W. R. & White, K. R. (1993). *Measurement and Evaluation in the Schools.* New York: Longman.
Yılmaz, H. (2004). *Eğitimde Ölçme ve Değerlendirme*, Konya: Çizgi Kitabevi Yayınları, 7th edition.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar