

THE EFFECT OF DIFFERENT TYPES OF TEST ON PRESERVICE CHEMISTRY TEACHERS' ACHIEVEMENT RELATED TO "CHEMICAL BONDING"

Senar Temel, Sinem Dinçol Özgür, Ayhan Yılmaz

Hacettepe University, Ankara, Turkey

E-mail: senar@hacettepe.edu.tr, sinemdincol@hacettepe.edu.tr,

ayhany@hacettepe.edu.tr

Abstract

The purpose of this study was to determine preservice chemistry teachers' achievement on different types of test and investigate the effect of different types of test on their achievement related to "Chemical Bonding". The participants of this study consisted of 26 preservice chemistry teachers in Hacettepe University, Faculty of Education, and Department of Chemistry Education in fall semester of 2010 – 2011 academic year. The mean age of preservice chemistry teachers was 20 years. In the study, Chemistry Achievement Tests that are designed to involve four different item types (Two-Tier Multiple-Choice, Multiple-Choice, Essay, Correct/Incorrect) towards the same behavioural objectives and are administered to the same group of students were used as data collection tools. The hypotheses were tested by using One Way ANOVA. A significant difference was found between preservice chemistry teachers' means of achievement scores on "Two-Tier Multiple-Choice Test", "Multiple-Choice Test", "Essay Test" and "Correct/Incorrect Test". The results showed that preservice chemistry teachers' are most successfully on "Multiple-Choice Test" and "Correct/Incorrect Test" and then "Two-Tier Multiple-Choice Test" and least successfully on "Essay Test".

Key words: achievement, chemical bonding, correct/incorrect test, essay test, multiple-choice test, two-tier multiple-choice test.

Introduction

In educational settings, tests are usually considered assessment tools (Roediger & Karpicke, 2006). The heavy emphasis on assessment often obscures another function of testing. This function is the promotion of learning and highly relevant to the goals of education (Butler & Roediger, 2007). A research has found that retention of studied material can be enhanced by testing (Kang, McDermott & Roediger, 2007). Also recent studies demonstrated that taking a test on studied material promotes learning and conceptual understanding (Cranney, Ahn, McKinnon, Morris & Watts, 2009; Roediger & Marsh, 2005). Roediger and Karpicke (2006) reported that testing is a powerful means by which to improve student learning rather than just assessing students' knowledge levels.

Different types of test are used to assess students' knowledge. Especially multiple-choice tests are used in educational settings but with unknown effects on student knowledge (Roediger & March, 2005). There are many advantages and also disadvantages to multiple-choice tests. They can be marked easily and are generally not time-consuming to prepare and administered to classes. They can produce objective scores for the purpose of assessment (Chang, Yeh & Barufaldi, 2010). Reading the choices in a test or retrieving information by cues might activate relative memory, modify the memory trace of target items and increase

the probability of a successful retrieval later (Kang, McDermott & Roediger, 2007). However multiple-choice tests consist of questions with several false statement choices and only one correct answer. The distracters in multiple choice tests expose students to a considerable amount of incorrect information. Reading the choices might increase the student's familiarity with the subject, but also cause them to misperceive erroneous information as correct, causing negative consequences later, especially among students who were less familiar with the subject area in the first place (Chang, Yeh & Barufaldi, 2010). Studies have found that distracters often lead to a negative effect and they are selected as a correct answer on future exams (Brown, Schilling, and Hockensmith 1999; Toppino and Luipersbeck 1993). Two-tier multiple-choice tests are used to determine students' alternative concepts. The first part of each item consists of a multiple-choice content question having usually two or three choices. The second part of each item consists a set of four or five possible reasons for the answer to the first part. So it allows an insight to the underlying reasons for students' answers (Treagust, 1995). Two-tier questions have two main benefits over conventional one-tiered questions. The first is a decrease in the measurement error. A two-tier question is considered correct only if both tiers are answered correctly. So, chance of correctly guessing the answer decreases. The second benefit is that it allows for the probing of two aspects of the same phenomenon. In the first tier, students are asked to predict the outcome of a chemical change and the second tier asks for an explanation. This allows the probing of the phenomenological domain with the first tier and the conceptual domain with the second (Tüysüz, 2009). True/false questions can be easier to prepare and quicker to read and answer, so that the number in a test is increased more easily and the subject matter is better sampled (Burton, 2001). Because random guessing will produce the correct answer half the time, true-false tests are less reliable than other types of exams (Davis, 1999). Chance may affect scores in multiple choice and true-false tests in two ways. First, if the questions sample only part of the examinable subject matter, then a particular examinee may be lucky or unlucky in the examiner's choice of questions (Posey, 1932). Second, marks may be obtained by guessing (Burton, 2001). Essay tests enable teachers to judge students' abilities to organize, integrate, interpret material and express themselves in their own words. It gives an opportunity to comment on student' progress, the quality of their thinking, the depth of their understanding and the difficulties they may be having. However the reliability of essay tests is compromised by subjectivity or inconsistencies in grading (Davis, 1999).

As mentioned above, different types of test have both advantages and disadvantages but they are used to assess students' knowledge without their effect on student knowledge. So it is important to know the positive or negative effect of different types of test on students' understanding and learning. When the literature are examined, it is seen that there are studies which focus on testing effect (how test format affects later memory performance), namely the role of tests in improvement of learning and promotion of permanent learning (Chang, Yeh& Barufaldi, 2010; Cranney, Ahn, McKinnon, Morris & Watts, 2009; Kang, McDermott & Roediger, 2007; McDaniel, Anderson, Derbish & Morrisette, 2007; Roediger & Karpicke, 2006) rastlanmaktadır. However, there are limited studies which examine student achievement on different types of test. Hence it was aimed to determine preservice chemistry teachers' achievement on different types of test and investigate the effect of different types of test on their achievement related to "Chemical Bonding".

Problem of Research

The purpose of this study was to determine preservice chemistry teachers' achievement on different types of test and investigate the effect of different types of test on their achievement related to "Chemical Bonding". In this aspect, the following two questions were tried to answer:

1. What are the achievement scores of the preservice chemistry teachers on "Two-Tier Multiple-Choice Test", "Multiple-Choice Test", "Essay Test" and "Correct/Incorrect Test"?

2. Is there a significant difference between achievement scores of preservice chemistry teachers on "Two-Tier Multiple-Choice Test", "Multiple-Choice Test", "Essay Test" and "Correct/Incorrect Test"?

Methodology of Research

Sample of Research

The participants of this study consisted of 26 preservice chemistry teachers in Hacettepe University, Faculty of Education, and Department of Chemistry Education in fall semester of 2010 – 2011 academic year. The mean age of preservice chemistry teachers was 20 years.

Instruments

In the study, Chemistry Achievement Tests that are designed to involve four different item types (Two-Tier Multiple-Choice, Multiple-Choice, Essay, Correct/Incorrect) towards the same behavioral objectives were used as data collection tools.

1) *The Two-Tier Multiple-Choice Test (TTMCT)*: The TTMCT was developed by Dinçol Özgür (2011). It consists of 15 two-tier multiple choice items related to the "Chemical Bonding". In TTMCT, students are asked, after checking the answer of the question, as a second step of the question, to check the reasons of their choices from again the given alternatives. The test was reviewed by the experts in the field of chemistry education after it had been prepared to ensure its' content validity. Cronbach's alpha reliability of the test was found to be 0.85.

Sample Question 1: Regarding chemical bonding, which among the below expression is true?

- a) When chemical bonds are formed, electron exchange always occurs.
- b) While energy is liberated during bond formation; in order for the bond to be ruptured energizing is required. This energy is called bonding energy. *
- c) Chemical bonds are formed through the sharing of the electrons in the valence shell by the atoms.

Which of the below expressions is the reason to your choice?

I) Chemical bond is formed through one of the atom's emitting electron and the other atom's capturing electron.

II) Bond formation between two atoms is an energy liberating situation, bond rupture is an energy requesting situation.*

III) Through the valence electrons in the valence shell being attracted by the nucleus of both two atoms, electrons are shared and chemical bond is formed.

IV) Bonding of the atoms to each other emerges through the atoms' transferring their electrons to each other.

2) *The Multiple Choice Test (MCT)*: The MCT was developed by Dinçol Özgür (2011). It consists of 15 multiple choice items related to the "Chemical Bonding". Experts in the field of chemistry education reviewed an initial version of the test regarding: (a) the adequacy of

the test's chemistry content with respect to students' developmental appropriateness, and (b) clarification and comprehensibility of the statements. Cronbach's alpha reliability of the test was found to be 0.82.

Sample Question 1: Regarding chemical bonding, which among the below expression is true?

- a) When chemical bonds are formed, electron exchange always occurs.
- b) While energy is liberated during bond formation; in order for the bond to be ruptured energizing is required. This energy is called bonding energy. *
- c) Chemical bonds are formed through the sharing of the electrons in the valence shell by the atoms.

3) *The Essay Test (ET):* The ET is composed of MCT questions being converted to open ended questions by the researchers. The questions are evaluated by experts in order to provide the content validity.

Sample Question 1: What is bonding energy, explain?

4) *The Correct/Incorrect Test (CIT):* The CIT including 15 items (7 correct/8 incorrect) related to "Chemical Bonding" prepared by the researchers. The choices in the MCT are used while preparing CIT. The students are requested to answer the test items as "true" or "false". Cronbach's alpha reliability of the test was found to be 0.80.

Sample Statement 1: While energy is liberated during bond formation; in order for the bond to be ruptured energizing is required. This energy is called bonding energy (Correct or Incorrect)

After teaching the topic of "Chemical Bonding", first the Essay Test, than the other tests are applied to the preservice chemistry teachers.

Data Analysis

In the study, the data analysis was carried through descriptive statistic (mean) and one way ANOVA. The level of significance is defined as 0.05 in data analysis.

Results of Research

Regarding the first question; the mean scores of achievement of the preservice chemistry teachers on 4 different kinds of tests are calculated. The results are given in Table 1.

Table 1. Mean scores of achievement of preservice chemistry teachers on different types of test

Test Types	Mean	N
1. The Two-Tier Multiple-Choice Test (TTMCT)	10.65	26
2. The Multiple Choice Test (MCT)	11.88	26
3. The Essay Test (ET)	6.99	26
4. The Correct/Incorrect Test (CIT)	11.88	26

When Table 1 is analyzed, it is determined that the TTMCT mean scores of achievement

of the preservice chemistry teachers are calculated as $\bar{X} = 10.65$, MCT mean scores of achievement as $\bar{X} = 11.88$, ET mean scores of achievement as $\bar{X} = 6.99$ and CIT mean scores of achievement as $\bar{X} = 11.88$. It is concluded that the test type in which the preservice chemistry teachers are the most successful are CIT and MCT ($\bar{X} = 11.88$), followed by TTMCT ($\bar{X} = 10.65$) and that the test type they are the least successful is ET ($\bar{X} = 6.99$). The achievement rates of the preservice chemistry teachers in CIT and MCT tests came out the same. In TTMCT, unlike MCT, the preservice chemistry teachers are asked, after checking the answer of the question, as a second step of the question, to check the reasons of their choices from again the given alternatives. That the success rates of preservice chemistry teachers in the TTMCT are lower shows that they might have checked this answer by chance without knowing the right answer they checked in the MCT. The test type in which they were least successful is ET. In this test type, it is expected from the preservice chemistry teachers themselves to answer the proposed questions by using their knowledge. That is, it is not in question for them to evaluate the given expressions as correct/incorrect as in CIT or to be directed in a way as to choosing the right answers from the given alternatives as in MCT and TTMCT.

Regarding the second question, one way ANOVA analysis is performed. The result of the one way ANOVA analysis is given in the Table 2.

Table 2. The ANOVA result of the TTMCT, MCT, ET and CIT mean scores of achievement of the preservice chemistry teachers.

Source	Sum of Squares	df	Mean Square	F	p
Between groups	381.882	25	15.275	77.41	0.000
Measure	417.853	3	139.284		
Error	134.939	75	1.799		
Total	934.674	103			

When Table 2 is analyzed, it is determined that there is a significant difference between the TTMCT, MCT, CIT and ET mean scores of achievement of the preservice chemistry teachers regarding the topic of "Chemical Bonding" ($F_{(3-75)} = 77.41$). It is ascertained that there is significant difference between TTMCT and CIT, TTMCT and ET, TTMCT and MCT mean scores of achievement and ET and CIT, ET and MCT mean scores of achievement. However, it is determined that there is no significant difference between the CIT and MCT mean scores of achievement in which the preservice chemistry teachers obtained the highest mean scores of achievement.

Discussion

In their study, Lawrenz, Huffman and Welch (2001) as well have analyzed student achievements by different assessment tools (multiple-choice test, a written open-ended test, a hands-on lab skill test and a hands-on full investigation) and reached to the conclusion that the least achievement rates are scored through the open ended questions.

In their study, Kızılcık and Tan (2007) could not find a significant relation between true-false test and short answer test although they have found significant relation between multiple-choice test and true-false test, and between short-answer test and multiple choice test applied on the issue of Repulsion and Momentum. They have reached the conclusion that when the assessment tools aiming at evaluating the same target behaviours, the evaluation results could be different as well. In their study, Lawrenz, Huffman and Welch (2001) have stated that different assessment tools evaluate different skills and that the achievement of the students in different subgroups shows a change in accordance with different assessment tools.

Conclusions

- The achievement scores scored by the preservice chemistry teachers from different test types applied differently yet on the same topic differ. In this context, since applying different assessment tools in evaluating the students would lead to difference in their achievement, different test types should be included in the evaluation process. In this way, enrichment and development of assessment process could be enabled.
- Moreover, each test type contains advantages and disadvantages within itself. Therefore, the characteristics of every test type in assessment procedures should be known and different types of tests should be prepared considering the characteristics of the topic and the student to be implemented on. Since particularly the learning characteristics the students have (learning styles, cognitive styles, etc.) could be effective on differentiation of their achievement in different test types, learning characteristics and assessment in different test types can be investigated in future studies.
- Also the effect of different types of test on achievement can be examined by studying larger sample in future studies.

References

- Brown, A. S., Schilling, H. E. H., & Hockensmith, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, 91 (4), 756-764.
- Burton, R. F. (2005). Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers. *Assessment & Evaluation in Higher Education*, 26 (1), 41-50.
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19 (4/5), 514-527.
- Chang, C., Yeh, T., & Barufaldi, J. (2010). The positive and negative effects of science concept tests on student conceptual understanding. *International Journal of Science Education*, 32 (2), 265-282.
- Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, 21 (6), 919-940.
- Davis, B. G. (1999). *Quizzes, tests and exams*. Retrieved 18/01/2012, from <http://www2.honolulu.hawaii.edu/facdev/guidebk/teachtip/quizzes.htm>.
- Dinçol Özgür, S. (2011). *Türetimci çoklu ortamın öğretmen adaylarının öğrenme stillerine göre başarı, tutum ve kalıcılığa etkisi ile öğretmen adaylarının ortama yönelik görüşleri*. Yüksek Lisans Tezi, Hacettepe Üniversitesi.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19, 528-558.
- Kızılcık, H. Ş., & Tan, M. (2007). Fizik öğretiminde kullanılan yazılı ölçme türlerinin itme- momentum konusu için karşılaştırılması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 27 (2), 109-122.

- Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education*, 85 (3), 179-290.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19 (4/5), 494-513.
- Posey, C. (1932). Luck and examination grades. *Journal of Engineering Education*, 60, 292-296.
- Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology-Learning Memory and Cognition*, 31 (5), 1115-1159.
- Roediger, H. L., & Karpicke, J. D. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255.
- Toppino, T. C., & Luipersbeck, S. M. (1993). Generality of the negative suggestion effect in objective tests. *The Journal of Educational Research*, 86 (6), 357-362.
- Treagust, D. F. (1995). *Diagnostic assessment of students' science knowledge*. In: Glynn SM, Duit R (Eds.), *Learning science in the schools: Research reforming practice*. Mahwah, New Jersey: Lawrence Erlbaum Associates. p. 327-346.
- Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essays*, 4(6), 626-631.

Advised by Vincentas Lamanaukas, University of Siauliai, Lithuania

Received: January 23, 2012

Accepted: March 06, 2012

Senar Temel	Dr., Expert, Hacettepe University, Ankara/ Turkey. E-mail: senar@hacettepe.edu.tr
Sinem Dinçol Özgür	PhD, Research Assistant, Hacettepe University, Ankara/ Turkey. Phone: + 0090 312 297 67 87. E-mail: sinemdincol@hacettepe.edu.tr
Ayhan Yılmaz	Prof. Dr., Hacettepe University, Ankara/ Turkey. E-mail: ayhany@hacettepe.edu.tr