Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context

Serpil KOÇDAR (Main Author)
Faculty of Open Education, Anadolu University, Turkey
skocdar@anadolu.edu.tr

Nejdet KARADAĞ
Anadolu University,Open Education Faculty, Yunus Emre Campus, 26470, Eskisehir/Turkey
nejdetkarada@gmail.com

Murat Doğan ŞAHIN
Faculty of Open Education, Anadolu University, Turkey
muratdogansahin@gmail.com

ABSTRACT
This is a descriptive study which intends to determine whether the difficulty and discrimination indices of the multiple-choice questions show differences according to cognitive levels of the Bloom's Taxonomy, which are used in the exams of the courses in a business administration bachelor's degree program offered through open and distance learning in a public university in Turkey, and to obtain the opinions of the learners on the cognitive levels of the questions. The study population consisted of 905 multiple questions which were asked in the mid-term, final, and make-up exams in the 11 major area courses. Quantitative data were gathered from item analysis reports. As well as that, qualitative data were obtained via semi-structured interviews with 20 learners. As a result, although some learners stated that they answered applying-level questions more easily, the learners were generally observed to answer the remembering and understanding-level questions more easily than the applying-level questions in parallel with the literature. Contrary to the studies in the literature, the remembering and understanding-level questions better distinguished the learners who received high scores from the learners who received low scores compared to the applying-level questions.

INTRODUCTION
Assessment of learning is an important element of an instructional design process, which provides feedback on learning and teaching processes and enables to review and improve the whole process (Haladyna, 2002). A variety of tools and techniques are used to assess learning in higher education such as assignments, tests, essays, portfolios, projects or oral examinations (Parker, 2005). One of the most common used tools has been the standardized achievement testing, which became popular in the early 1920s in the United States after the emergence of mass education (Haladyna, 2002). The use of standardized achievement tests consisting of multiple-choice questions is widespread as they are practical and provides objective results especially for mega universities with large number of learners in open and distance learning, in which learners, teachers, and learning sources are not in a central location (Simonson et al., 2012; Zhang, 2002).

Multiple-choice tests are analyzed through various methods and new tests are developed based on the outcomes of the analyses. One method is the item (question) analysis which is a process that examines learner responses to individual test items in order to assess the quality of those items and of the test as a whole. The *difficulty* (p) and *discrimination* (r) *indices* of the items are calculated in this analysis (Özçelik, 1989). Item difficulty is the percentage of learners who answered an item correctly and ranges from 0.0 to 1.0. The closer the difficulty of an item approaches to zero, the more difficult that item is. The discrimination index of an item is the ability to distinguish high and low scoring learners. The closer this value is to 1, the better the item distinguishes the learners who get a high score from those who get a low score. Analysis of each item by calculating difficulty and discrimination indices provides feedback on what the learners have learned and enables instructors to determine and correct the faulty items. In other words, it contributes to increasing the validity and reliability of the tests by revealing whether the items are working well or not.

Multiple-choice tests are prepared according to learning taxonomies. There are many taxonomies in the literature (Anderson & Krathwohl, 2001; Biggs and Collis, 1982; Bloom, 1956; Fink, 2003; Hannah & Michaelis, 1977; Marzano, 2001; Stahl & Murphy, 1981). The most commonly used taxonomy is Bloom's taxonomy of cognitive

domain (Haladyna, 2002; Seaman, 2011). According to first version of Bloom's taxonomy, there are six categories of cognitive domain which are *knowledge, comprehension, application, analysis, synthesis,* and *evaluation*. The categories proceed in a hierarchical structure, from simple to complex. Bloom's taxonomy has been updated in line with the developments in cognitive psychology and learning. *Knowledge* has been replaced with *remembering*, *comprehension* has been replaced with *understanding*, and the highest level cognitive step is determined as *creating* in the new taxonomy (Krathwohl, 2002). The categories are *remembering, understanding, applying, analyzing, evaluating*, and *creating* in the new Bloom's taxonomy.

The literature includes many studies on analyzing exam questions according to cognitive levels. These studies mainly deal with which cognitive domain category the exam questions fall into or the relationship between the difficulty and discrimination indices (Demircioğlu & Demircioğlu, 2009; Gümüş et al., 2009; Hingorjo & Jaleel, 2012; Pande et al., 2013; Sim & Rasiah, 2006; Tanık & Saraçoğlu, 2011). On the other hand, there are a limited number of studies on the relationship between cognitive levels and difficulty and discrimination indices of exam questions. These studies show that the effect of cognitive levels on the difficulty and discrimination indices of the questions are not parallel; the results differ according to the subject and context. For example, Momsen et al. (2013) found no relationship between the difficulty and cognitive levels (according to Bloom's taxonomy) of the questions for a biology course, and a poor relationship for the questions of a physics course in their study conducted at the bachelor's level. On the other hand, Veeravagu, Muthusamy, Marimuthu, and Michael (2010) found a relationship between the cognitive levels in Bloom's taxonomy and the performance of the learners for the questions of an English reading skills course. According to the researchers, the learners had difficulty in answering specifically the questions of high-level cognitive skills: analysis, synthesis, and evaluation. In parallel, Nevid and McClelland (2013) indicated that the learners had difficulty in answering the questions of *evaluation* and *explanation* at high cognitive levels in Bloom's taxonomy for a psychology course, and these kinds of questions were the most distinctive for high-performing and low-performing learners. In another study, Kim et al. (2012) found the difficulty indices of the multiple-choice questions in pharmacy studies at the *remembering*, *understanding*, and *applying* levels to be higher than the questions at the *analysis* and *synthesis/evaluation* levels. However, the discrimination indices of the questions at the *application* and *synthesis/evaluation* levels were higher than the questions at *remembering* and *understanding* levels.

In this regard, this study aims to determine whether the difficulty and discrimination indices of the multiple-choice questions show differences according to cognitive levels, which are asked in the exams of the courses in a business administration bachelor's degree program offered through open and distance learning in a public university in Turkey, and to obtain the opinions of the learners on the cognitive levels of the questions. No studies were found in the literature on the questions of business administration programs which is one of the most common programs in higher education in the world that includes a large number of learners. Research questions are as follows:

1. Do the difficulty indices (p) of multiple-choice questions show a significant difference according to cognitive levels?
2. Do the discrimination indices (r) of multiple-choice questions show a significant difference according to cognitive levels?
3. What are the learners' opinions about the questions asked at different cognitive levels?

METHOD

This is a descriptive study which intends to investigate whether the difficulty and discrimination indices of multiple-choice questions differ according to cognitive levels in a business administration program offered through open and distance learning.

Study population and the participants

The study population consisted of 905 multiple questions (with 5 choices) which were asked in the mid-term, final, and make-up exams in the 2011–2012 fall and spring semesters in the 11 major area courses of a business administration bachelor's degree program at a public university in Turkey. No sampling was made; all of the questions in the population were used. The questions of the business administration program were selected because this department has the largest number of learners in the university with about 350,000 learners.

The participants of the study consisted of 20 volunteer learners in the Department of Business Administration. The learners were selected using a convenience sampling method. The demographic characteristics of the learners are shown in Table 1. The learners were coded as L1, L2, L3 and so on to keep their identity confidential.

Table 1. Demographic Information of Learners

Learners	Gender	Age
L1	Female	40
L2	Female	22
L3	Male	23
L4	Female	22
L5	Male	27
L6	Male	30
L7	Male	23
L8	Female	38
L9	Female	21
L10	Female	23
L11	Female	22
L12	Female	29
L13	Male	27
L14	Female	40
L15	Female	27
L16	Female	21
L17	Female	32
L18	Male	26
L19	Female	27
L20	Male	35

Data collection tools

Quantitative data were collected for the first and second research questions, and qualitative data were collected for the third research question.

Quantitative data collection tools

The *item analysis documents* prepared for each course, which are prepared by the Information Processing Department of the university with the use of computer programs after each exam, were used to determine the difficulty (p) and discrimination (r) indices of the 905 questions in the study. Item analysis documents are prepared by comparing the answers of the group scoring the highest 27% and the group scoring the lowest 27% to each item after putting the scores in an order from high to low in a test. To analyze the items, first the questions are graded and the number of true answers of the learners are counted for the entire test; the number of true answers are taken as the score. After scoring is completed, the answer sheets are put in order from the highest to the lowest with the paper with the highest score placed on the top. Then the answers from the top and bottom 27% scored papers are analyzed (Özçelik, 1989). The lowest and highest numbers of learners who took the exams for the courses for which item analysis was performed were 1,998 and 71,210, respectively.

Qualitative data collection tools

The qualitative data were collected through semi-structured interviews. The interview questions were corrected in line with the opinions of three experts after being formed.

Data Collection

Quantitative data collection

Bloom's revised cognitive domain taxonomy that includes categories of *remembering, understanding, applying, analyzing, evaluating,* and *creating* was used in determining the cognitive levels of the questions in this study because it had been commonly used in the literature (Seaman 2011).

In the first step, the cognitive levels of the questions were coded by three assessment experts, and the inter-coder reliability was calculated using the formula (Inter-coder reliability= Agreement / Agreement + Disagreement) of Miles and Huberman (1994) and found to be 95%. Coders had disagreement on 39 of the 905 questions. So, they reviewed the 39 questions together on which disagreement occurred and reached an agreement on the cognitive levels of these questions. The questions were observed to be distributed at the first three levels, *remembering, understanding,* and *applying*, of Bloom's taxonomy. The distribution of the questions according to cognitive levels are shown in Table 2.

Table 2. The Distribution of the Questions according to Cognitive Levels

Cognitive Levels	**Number of Questions**	**Percentage (%)**
Remembering	350	38,6
Understanding	474	52,4
Applying	81	9,0
Total	**905**	**100,0**

After determining the cognitive levels of the questions, the p and r indices of each item was identified from item analysis documents and tabulated to be analyzed.

Qualitative data collection

The learners in the Department of Business Administration were accessed through phone and social media for semi-structured individual interviews and were informed of the subject and scope of the study. It was explained to the participants that their identities would be kept confidential and would not be shared with third parties. The learners who volunteered to participate in the study were interviewed through Skype or phone on the determined date. The permission of the learners were obtained to record the interview.

Data Analysis

Quantitative data analysis

Data were analyzed by SPSS program. One-way MANOVA Test was used. When a significant difference was found in One-way MANOVA results, the One-way ANOVA was used to determine the dependent variables that caused the difference. When a significant difference was found as a result of One-way ANOVA, Scheffe was used in cases where the homogeneity of variances assumption was ensured, and the Brown-Forsythe and Welch Test was used in cases where the homogeneity of variances was not ensured. Pairwise comparisons were made using Tamhane's T2 tests if significant results were found.

The assumptions required for MANOVA had to be checked to determine whether the difficulty and discrimination indices of the questions differed according to cognitive level using the One-way MANOVA. In addition to its advantages of testing multiple dependent variables at once (Field, 2005) and protecting against Type I errors (Bray & Maxwell, 1982; Stevens, 2009; Stangor, 2010), MANOVA also brings forth many assumptions. Checking the assumptions of univariate and multivariate normality, outliers, linearity, multicollinearity and singularity, and homogeneity of covariance matrices are the prerequisites to apply MANOVA (Pallant, 2005). Therefore, these mentioned assumptions were checked before the One-way MANOVA analyses.

At first, univariate normality of dependent variables was checked by the Kolmogorov-Smirnov (K-S) Test and the results were found to be statistically significant ($p<0.01$). However, the results of K-S should not be found to be statistically significant to meet the assumption of univariate normality. Nevertheless, it is known that even much smaller deviations can be found as significant when a large number of data is present in the study (Çetin, İlhan, & Arslan 2012). Considering that the coefficient of skewness of the data between ± 1 can be interpreted as the scores do not show a significant deviation from normal (Büyüköztürk, 2010), the coefficient of skewness of the dependent variables were analyzed to make the final decision. The skewness values were found to be -0.019 and 0.054 for the difficulty and discrimination indices of the questions, respectively. Thus, it was found that the dependent variables met the univariate normality condition.

After meeting the univariate normality condition, Mahalanobis distance values were calculated to test whether or not the data met the multivariate normality assumption. Pearson, Pearson and Hartley (1958) reported the critical value for Mahalanobis distance to be 13.82 in a multivariate analysis with two independent variables. The Mahalanobis values above this critical value are accepted as extreme values (Pallant, 2005). In this study, three Mahalanobis values (14.77, 14.12, and 13.95) were above the critical value of 13.82. These extreme values were

considered to negatively affect the results of the study and therefore, they were deleted. The study continued with the data set of 902 multiple-choice questions.

Another assumption to be checked before MANOVA is to determine whether or not a linear relation exists among the dependent variables. The relationship among the dependent variables should be linear in all categories of independent variables. The graphics obtained on the linearity of all paired combinations of the dependent variables (difficulty and discrimination indices) in all categories of cognitive levels (remembering, understanding, and applying) of the questions.

MANOVA provides the best results when a medium-level correlation exists among the dependent variables. Univariate variance analysis should be applied when the correlation is low. Over 0.80 or 0.90 correlation among the dependent variables means multicollinearity and causes problems in MANOVA (Pallant, 2005). In this study, the correlation analysis showed a medium-level relation of 0.49 among the dependent variables, so no multicollinearity occurred.

Finally, Box's M Test was used to check the assumption of the homogeneity of variance-covariance matrices. The results of the Box's M Test were found to be statistically significant, meaning that the homogeneity of variance-covariance matrices assumption was violated. In cases where numbers are not equal in the categories of the variables and Box's M Test reveals statistically significant results at $p<0.001$, robustness cannot be ensured. However, it should be noted that the Box's M Test may reveal statistically significant results due to extremely small changes in large sampling groups. In such cases, it will be appropriate to use Pillai's Trace as the evaluation criteria instead of Wilk's Lambda, which is generally used in MANOVA (Tabachnick & Fidell 2007). In this study, the results of Pillai's Trace Test showed that there were no situations that prevent to use MANOVA. In sum, the analyses showed that the dependent variable set consisting of the difficulty and discrimination indices of the questions met all the assumptions to apply One-way MANOVA to test the questions in terms of the independent variable of cognitive level.

Qualitative data analysis

The interviews were recorded, decoded, and analyzed using the descriptive analysis method. Yıldırım and Şimşek (2008) stated that descriptive analysis is more superficial than content analysis and is used in studies where the conceptual structure of the study is clearly previously determined. The data can be organized according to the themes set by the study questions or by the questions used during the interviews and observations. In this respect, the data were summarized and interpreted according to the interview questions. Two researchers coded the data for the reliability in data analysis. The inter-coder reliability was found to be at 85%. Agreement was ensured by discussing on the items that were coded differently.

FINDINGS

The descriptive statistics of the difficulty and discrimination indices of items according to cognitive levels were obtained first. These statistics are shown in Table 3.

Table 3: The Descriptive Statistics of the Items Categorized according to the Independent Variable

Independent Variable	Category	Number of Items	Dependent Variable	Mean	St. Deviation
Cognitive Level	**Remembering**	348	$\mathbf{p_{jx}}$	.464	.174
			$\mathbf{r_{jx}}$	.375	.155
	Understanding	473	$\mathbf{p_{jx}}$	.481	.183
			$\mathbf{r_{jx}}$	.359	.148
	Applying	81	$\mathbf{p_{jx}}$	.376	.146
			$\mathbf{r_{jx}}$	.326	.109
Total		**902**			

The mean scores of difficulty and discrimination indices of the 348 items at the remembering level were found to be 0.464 and 0.375, respectively. The standard deviations of these two values were 0.174 and 0.155, respectively. The mean scores for difficulty and discrimination indices of the questions at the understanding level were found to be 0.481 and 0.359, respectively. The standard deviations of these two values were 0.183 and 0.148, respectively. The mean scores of the difficulty and discrimination indices of the questions at the applying level were found to be 0.376 and 0.326, respectively. The standard deviations of these two values were

0.146 and 0.109, respectively. These findings suggest that the items with the highest difficulty index (the easiest items) were found to be the understanding-level items, and the items with the highest discrimination value were found to be the remembering-level items. One-way MANOVA Test results are shown in Table 4.

Table 4: MANOVA Test Results according to Cognitive Level

Test	F	Sig.
Pillai's Trace	7.663	.000*

p<.001

The Pillai's Trace coefficient was found to be statistically significant ($p < 0.01$). This showed that the difficulty and discrimination indices of items were significantly different between at least two categories of the cognitive level of the item, which is the independent variable. One-way ANOVA should be applied to each dependent variable to determine which categories show differences, and the homogeneity of variances assumption should be met to apply One-way ANOVA. The Levene Test was carried out for this purpose and its results are shown in Table 5.

Table 5: Levene Test Results for the Independent Variable of Cognitive Level

	F	Sd	Sig.
Difficulty index	6.163	2	.002*
Discrimination index	6.838	2	.001*

*p<.05

The values obtained from the Levene Test were found to be statistically significant for the difficulty and discrimination variables ($p < 0.05$), so homogeneity of variances assumption could not be met. Therefore, Welch and Brown-Forsythe Tests were used before making paired comparisons for the levels in the independent variable using One-way ANOVA. The results are shown in Table 6.

Table 6: Welch and B-F Test Results for Difficulty and Discrimination Indices of the Items

Independent Variable		Sd	Sig.
Difficulty Index	**Welch**	2	.000*
	Brown-Forsythe	2	.000*
Discrimination Index	**Welch**	2	.004*
	Brown-Forsythe	2	.009*

*p<.025

Two separate analyses are used to analyze the effect of the same independent variable for the independent variables of the difficulty and discrimination indices of items. In these cases, the Bonferroni correction should be applied to prevent Type I error (Pallant, 2005). The easiest calculation of the Bonferroni correction is to divide the alpha (the generally used value is 0.05) into the number of dependent variables (Tabachnick & Fidell 2007). Since two dependent variables are in this study, the 0.05 alpha value was divided into 2 and the new alpha value was found to be 0.025 (0.05/2 = 0.025). Accordingly, the paired comparisons of the difficulty and discrimination indices of items for the categories of independent variables were found to be statistically significant (p<.025).

The next step was the Post-Hoc Tests for the paired comparisons of the mean scores of dependent variables according to the levels in independent variables because the Welch and Brown-Forsythe test results were statistically significant. Tamhane's T2 Test, one of the Post-Hoc Tests, was used since the homogeneity of variance assumption could not be met. The results are shown in Table 7.

Table 7: The Results of Tamhane's T2 Multiple Comparison Test for the Difficulty and Discrimination of Items according to Cognitive Level

Dependent Variable	Cognitive Level (I)	Cognitive Level (J)	Mean Difference (I-J)	Std. Error	Sig.
Difficulty Index	**Remembering**	**Understanding**	-.018	.013	.410
		Applying	.088	.019	.000*
	Understanding	**Remembering**	.018	.013	.410
		Applying	.106	.018	.000*
	Applying	**Remembering**	-.088	.019	.000*
		Understanding	-.105	.018	.000*
		Understanding	.016	.011	.378

Discrimination Index	**Remembering**	**Applying**	.049	.015	.003*
		Remembering	-.016	.011	.378
	Understanding	**Applying**	.034	.014	.048*
		Remembering	-.049	.015	.003*
	Applying	**Understanding**	-.034	.014	.048*

According to the results of Tamhane's T2 Test, no significant difference was found between the discrimination indices of the remembering and understanding-level questions
($p = 0.378 > 0.05$), and a significant difference was found between the discrimination indices of the questions at the remembering and applying levels ($p = 0.003 < 0.05$) and the questions at the understanding and applying levels ($p=.048<.05$).

The opinions of the learners on the cognitive levels of the questions showed differences in the individual interviews. Many learners expressed that they answered the remembering-level questions more easily and quickly, and they had difficulty in understanding and applying-level questions. On the other hand, L4 stated that she found the remembering-level questions to be more difficult and preferred the understanding and applying-levels questions. Similarly, L6 stated that he found the remembering and understanding-level questions to be more difficult and the applying questions to be easier. Examples of the learners' opinions are as follows:

L11: I have difficulty in understanding-level questions. And I most easily answer the remembering-level questions.

L13: I only answer the remembering-level questions more easily and quickly. The understanding and applying-level questions are both time-consuming and difficult...

L4: I think the remembering-level questions are more difficult. It is very hard to exactly remember the information in the book. Instead, as a hard-working learner, I prefer the understanding and applying-level questions. These questions better distinguish the hard-working learners.

L6: I find the remembering and understanding-level questions to be more difficult and the applying-level questions to be easier.

While some of the learners expressed that the questions measuring the high-level cognitive skills should not be asked, some other gave a positive opinion for questions in higher levels which are *analyzing, evaluating*, and *creating*. Examples of the learners' opinions are as follows:

L13: The commentary questions would not be useful for the Open Education Faculty; it would become harder to pass.

L6: Existence of high cognitive level questions would be challenging, therefore it would be useful for the hardworking learners.

L16: It would be difficult; I don't prefer.

L19: Sometimes there are such questions. I prefer commentary questions. I prefer and more easily answer the questions which are not exactly the same in the books. They make me think.

DISCUSSION AND CONCLUSION

This study aimed to determine whether the difficulty and discrimination indices of the multiple-choice questions in the exams of the courses in a business administration bachelor's degree program offered via open and distance learning showed differences according to cognitive levels, and to obtain the opinions of the distance learners on the cognitive levels of the questions. The questions in the study were found to be at three levels: remembering, understanding, and applying. Although some learners stated that they answered applying-level questions more easily, the learners were generally observed to answer the remembering and understanding-level questions more easily than the applying-level questions in parallel with the studies of different researchers in the literature (Kim et al., 2012; Nevid & McClelland, 2013; Veeravagu et al., 2010). The different opinions of the learners can be explained by the differences in their cognitive competencies.

The studies in the literature showed that the questions measuring high-level cognitive skills better distinguish the high-performing and low-performing learners compared to the questions measuring low-level cognitive skills (Kim et al., 2012; Nevid & McClelland, 2013). Contrary to the studies in the literature, the remembering and understanding-level questions better distinguished the learners who received high scores from the learners who received low scores compared to the applying-level questions in this study. In other words, the questions measuring low-level cognitive skills performed better. One reason for this may be the different subject- business administration- and the context of the study-the open and distance learning involving heterogeneous learner groups unlike the studies in the literature. The learners were mostly adults and varied in terms of formal

education, age, experiences, and characteristics when compared with the studies conducted in the context of traditional education in the literature.

Another reason that the questions measuring low-level cognitive skills performed better may be that the distracters (incorrect answers) were not strong enough and therefore the low-performing learners estimated the answers correctly even though they did not know the answers. Incorporation of strong distracters is crucial while forming questions to prevent this situation. In this regard, the analysis of the performance of the distracters could have included in the study which could lead to more correct interpretations. So, it is recommended to include distracter analysis for similar future studies.

The results of this study may be used as a guide but cannot represent all business administration programs. The study should be repeated for different sets of questions asked in the exams of the current business administration program in different years, and different business administration programs specifically offered via open and distance learning. Moreover, the question set in this study included remembering, understanding, and applying levels. A question set should be analyzed including the questions that also measure higher level cognitive skills to better explain the relationships between the cognitive levels and the difficulty and discrimination indices of the questions.

This study was conducted within the scope of Classical Test Theory, in which item parameters are dependent on the group. Scaling and analyzing the questions according to Item Response Theory may reveal different results. Therefore, the cognitive levels of the questions may be analyzed within the framework of Item Response Theory in future studies.

The questions which were examined in the study should have been prepared for the same learning outcome to ensure the consistency of the subject. However, the number of questions for a specific learning outcome was insufficient to make an analysis. Therefore, only the questions asked in the major area courses were included instead of the questions asked in all courses to ensure subject consistency.

The study is intended to contribute to the quality of assessment practices, and guide teachers, test developers and assessment experts while preparing multiple-choice questions. Assessment of learning is one of the most important elements of learning design process as it provides feedback to the learners and teachers and enables to improve the quality of the system. The validity and reliability of the assessment systems is one of the prerequisites for quality assurance and accreditation of institutions. However, especially in open and distance learning programs where there are a large number of learners, multiple-choice questions may be the only assessment tool. In this respect, conducting studies in different subject areas and contexts to determine the relationship between the cognitive levels and the difficulty and discrimination values of such questions is important to ensure validity and reliability of assessment tools and to increase the quality of questions in an open and distance learning context.

Authors' Note: This study was supported by Anadolu University Scientific Research Projects Commission under the grant no: 1406E308.

REFERENCES

Anderson, L. W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching and assessing: a revision of bloom's taxonomy of educational objectives.* New York: Longman.

Bloom, B. S. (1956). *Taxonomy of educational objectives: the classification of educational goals.* New York: Longmans, Gren and Company Inc.

Biggs, J. B., & Collis, K. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York, Academic Press.

Bray, J.H., & Maxwell, S.E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational Research,* 52, 340–367.

Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2011). *Sosyal bilimler için istatistik.* Ankara: Pegem Akademi.

Çetin, B., İlhan, M., & Arslan, S. (2012). Öğretmen adaylarının benimsedikleri eğitim felsefelerinin çeşitli değişkenler açısından incelenmesi. *The Journal of Academic Social Science Studies,* 5, 149-170.

Demircioğlu, G., & Demircioğlu, H. (2009). An evaluation of the questions chemistry teachers asked in exams in terms of the target behaviors. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education,* 3(1), 80-98.

Field, A. (2009). *Discovering statics using SPSS*. London: SAGE Publications Ltd.

Fink, L. D. (2003). *Creating significant learning experiences: an integrated approach to designing college courses*. San Francisco: Jossey-Bass.

Gümüş, İ., Ermurat, D.G., Kaya, Y., Kırıcı, M., & Kurt, M. (2009). Analysis for high school biology exam questions according to their cognitive development levels. *Erzincan Eğitim Fakültesi Dergisi,* 11(2).
Haladyna, T.M. (2002). *Standardized achievement testing*. Boston, MA: Allyn and Bacon.
Hannah, L. S., & Michaelis, J. U. (1977). *A comprehensive framework for instructional objectives: a guide to systematic planning and evaluation.* Reading, Mass: Addison- Wesley.
Hingorjo, M.R., & Jaleel, F. (2012). Analysis of one-best mcqs: the difficulty index, discrimination index and distractor efficiency. *Journal of Pakistan Medical Association,* 62(2), 142-157.
Kim, M. K., Patel, R.A., Uchizono, J.A., & Beck, L. (2012). Incorporation of Bloom's taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. *American Journal of Pharmaceutical Education,* 76 (6), 1-8.
Krathwohl, D. R. (2002). A revision of Bloom's taxonomy. *Theory into Practice.* 41(4), 212-218.
Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives.* Thousand Oak, California: Corwin Press, Inc.
Miles, M. B., & Huberman, A.M. (1994). *Qualitative data analysis: an expanded sourcebook.* 2nd ed. Thousand Oaks, California: SAGE.
Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE-Life Sciences Education*, 12, 239–249.
Nevid, J.S., & McClelland, N. (2013). Using Action Verbs as Learning Outcomes: Applying Bloom's taxonomy in measuring instructional objectives in introductory psychology. *Journal of Education and Training Studies,* 1(2), 19-24.
Özçelik, D.A. (1989). *Test hazırlama kılavuzu*. Ankara: ÖSYM Yayınları 8.
Pallant, J. (2005). *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows*. Australia: Allen & Unwin.
Pearson, E. S., Pearson, K., & Hartley, H.O. (1958). *Biometrika tables for statisticians*. New York: Cambridge University Press.
Pande, S. S., Pande, S.R., Parate, V.R., Nikam, A.P., & Angrekar, S. (2013). Correlation between difficulty & discrimination indices of MCQs in formative exam in physiology. *South-East Asian Journal of Medical Education,* 7(1), 45-50.
Parker, R. L. (2005). An investigation of assessment techniques used in American distance education. Unpublished PhD Dissertation, University of Cincinnati.
Seaman, M. (2011). Bloom's taxonomy: its evolution, revision, and use in the field of education. *Curriculum and Teaching Dialogue,* 13(1& 2), 29–43.
Sim, S.M., & Rasiah, R.I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Ann Acad Med Singapore,* 35(2), 67-71.
Simonson, M., Smaldino, S., Albright, M., & Zvacek, S. (2012). *Teaching and learning at a distance: foundations of distance education.* 5th ed. Boston: Allyn & Bacon.
Stahl, R. J., & Murphy, G.T. (1981). The domain of cognition: an alternative Bloom's cognitive domain within the framework of an information processing model. *ERIC Documents Reproduction Service* No: ED 208 511.
Stevens, J.P. (2009). *Applied multivariate statistics for the social sciences*. New York: Routledge.
Strangor, C. (2010). *Research methods for the behavioral sciences*. Boston, MA: Houghton Mifflin.
Tabachnick, B. G., & Fidell, L.S. (2007). *Using multivariate statics.* Boston: Pearson.
Tanık, N., & Saraçoğlu, S. (2011). Fen ve teknoloji dersi yazılı sorularının yenilenmiş Bloom taksonomisi'ne göre incelenmesi. *TUBAV Bilim Dergisi,* 4(4), 235-246.
Veeravagu, J., Muthusamy, C., Marimuthu, R., & Michael., A.S. (2010). Using Bloom's taxonomy to gauge students' reading comprehension performance. *Canadian Social Science,* 6(3), 205-212.
Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri.* Ankara: Seçkin.
Zhang, W., Tsui, C., Jedege, O., Ng, F., & Kowk, L. (2002). A comparison of distance education in selected Asian open universities. *Paper presented at 14th Annual Conference of Asian Association of Open Universities*, Manila, Philippines.