



THE PREDICTION OF ITEM PARAMETERS BASED ON CLASSICAL TEST THEORY AND LATENT TRAIT THEORY

MADDE PARAMETRELERİNİN KLASİK VE ÖRTÜK ÖZELLİKLER TEST TEORİLERİNE GÖRE KESTİRİLMESİ

Duygu ANIL*

ABSTRACT: In this study, the prediction power of the item characteristics based on the experts' predictions on conditions try-out practices cannot be applied was examined for item characteristics computed depending on classical test theory and two-parameters logistic model of latent trait theory. The study was carried out on 9914 randomly selected students who took the student selection and placement examinations applied in Turkey that is one the examinations on which try-out practices cannot be applied. In this study, the experts' prediction values were obtained from differently prepared metric scale applications according to classical and latent trait theories which were applied to 16 math teachers. In the study findings it was found that the item difficulty indexes of mathematic sub-test of student selection and placement examinations based on experts' predictions have the capability of predicting the item difficulty indexes obtained from the classical test theory. Furthermore the predictions of bi parameters based on the experts' prediction have the capability of predicting bi parameters based on logistic model with two parameters of latent trait theory. But it is also found out that item discriminating indexes obtained from the classical test theory and ai parameters obtained from latent trait theory cannot be estimated through the experts' predictions.

Keywords: classical test theory, latent trait theory, prediction of item and test parameters, test information functions

ÖZET: Araştırmada, deneme uygulaması yapılamayan durumlarda madde özelliklerinin uzman tahminlerine dayalı kestirimlerinin klasik test teorisi ve örtük özellikler teorisinin iki parametrelili logistik modeline dayalı olarak hesaplanan madde özelliklerini tahmin etme gücünün nasıl olduğu incelenmeye çalışılmıştır. Araştırma; Türkiye'de uygulanan deneme uygulaması yapılamayan sınavlardan biri olan Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavına giren random olarak seçilmiş 9914 öğrenci üzerinde yürütülmüştür. Araştırmada, uzman tahmin değerleri; 16 matematik öğretmenine uygulanan klasik ve örtük özellikler teorilerine göre ayrı ayrı hazırlanmış metrik ölçek uygulamalarından elde edilmiştir.

Araştırma bulgularında, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait madde güçlük indekslerinin uzman tahminlerinin klasik test teorisine dayalı madde güçlük indekslerini ve uzman tahminlerine dayalı bi parametre kestirimlerinin örtük özellikler test teorisinin iki parametrelili logistik modeline dayalı bi parametrelerini yordayabildiği görülmüştür. Fakat, klasik test teorisinden elde edilen madde ayırıcılık gücü indeksi ve örtük özellikler teorisinden elde edilen ai parametresinin uzman tahminleri ile kestirilemeyeceği görülmüştür.

Anahtar sözcükler: klasik test teorisi, örtük özellikler teorisini, madde ve test parametrelerinin tahmini, test bilgifonksiyon

1. INTRODUCTION

Measurement reserves an important place in both scientific studies and in daily works. In general terms, measurement is the observation of any characteristic and expression of those observation results with numbers or other symbols. (Turgut, 1997). Today, sciences develop themselves as much as they develop their measurement techniques and as well as they can make their measurement techniques available to measure even very small units. The concept of "without measurement there won't be any sciences" is the way of expressing reality. Thorndike (1971) who believes "if something exists in nature, it exists in definite quantity and Mc Call who says "if something exists in definite quantity, it should be measurable" both emphasize that measurement is necessary for all fields of science and it is certain that there is no need for any argument.

In behavioral sciences, which include education and psychology, qualities that we would like to measure are not the variables we can observe directly as in physical sciences.. Hence, examination of the truthfulness of theoretical relevance in social sciences may be possible as far as based on observation and measurement sensitiveness. Therefore for the definition of variables in social sciences

* Öğr.Gör.Dr., Hacettepe Üniversitesi, aduygu@hacettepe.edu.tr

with operational techniques development of measurement techniques and tools is essential so as to improve the reliability of datum.

Among the measurement techniques that are applied in education which multiple-choice tests are superior to others for both reliability and validity. In measurements that are used in education and psychology with the development of multiple-choice test, which features are estimated previously may be achievable. As Baykul (2000) has also stated, in obtaining a test with the required elements, after the item parameters are predicted from the data provided by the try-out group, with the benefit of the mathematical relationship between item and test parameters, items that provide the wanted test parameters are selected.

Today, the most commonly practiced theory, which is used in item analysis studies, is the classical test theory. In classical test theory, determination of psychological characteristic degrees that an individual owns is acquired by taking total reactions of each individual to the measurement tool item that is prepared in order to measure that characteristic. In other words, row score of an individual that she/he has obtained from the measurement tool is the indicator of the degree, which that person possesses (Fan, 1998). In addition, after 1950s, latent trait theory has also been effective in the development of measurement tools. This theory (LTT) as well as item analysis gives opportunity to register individuals based on their characteristics on a scale. While latent trait theory takes determination of ability as a goal, item characteristic curve model forms the core program of this theory. Since 1960s there have been significant improvements in item analysis and in order to have better predictions new item curve models have been developed (Baker, 1977).

In latent trait theory, related to mathematical structures of item characteristic functions different models have come to existence (Stocking, 1997). In the usage of these models it is essential to provide assumptions of latent trait theory. These assumptions include the distribution being normal, one-dimensional and having local independence.

Although models based on latent trait theory are developed multi-dimensional, the theory assumes that the measured item pool is one dimensional. Furthermore, all one-dimensional models in this theory have the following assumption: only one latent structure (θ) is the product of predicted reactions that are given to test item and reactions of (θ) to an item may change due to its way of forming (Harvey and Hammer, 1999).

The above mentioned differences have led to a variety of latent trait theory models. There of those are three most frequently used latent trait theory models for two-category scales. This study used the two-parameter logistic model, which is one of these three models. These models define the formation of the relationship between the item parameters and item reactions that are observed by θ in every situation.

1.1. Two-Parameter Logistic Model

$P^i(\theta)$, b_i , a_i , (θ) values used in this model. Have the same meaning with the ones that are used in the normal ogive model. This model includes D scale of measurement function. The difference between two-parameter logistic model and two-parameter normal ogive model is that when a_i is multiplied with the D value that is 1.7, it gets closer to normal ogive model (Hambelton and Swaminathan, 1985). Item characteristic function of this model is given below:

$$P^i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad (i= 1,2,3\dots n) \quad (1)$$

(Hambelton and Swaminathan, 1985, p.49).

b_i parameter of an item is the probability of the item's being correctly answered by 0.50, which refers to a value on the θ -ability-scale. b_i parameter demonstrates the degree of latent trait or the degree of item difficulty and in theory it is said that it gets value around $(-\infty, +\infty)$. Hambelton and Swaminathan (1985) stated that when a group scores are transferred as $X=0$ and $S_x=1$, this parameter

gets a value around (-2;+2) in practice. As the item becomes simpler this value gets closer to -2 and as it becomes more complex it is closer to +2. Items whose item difficulty indexes, which are around 0, are the items at a moderate difficulty. This parameter is also called the ability scale that an item measures.

Two-parameter logistic model adds discriminating power or a new parameter that we call a_i to the model. And this parameter let item characteristic curves have different slopes for different items. The parameter of discriminating power shows that some items have more relation or less relation with θ compared to others. This parameter is important as the amount of information is obtained directly from a_i parameter (Harvey and Hammer, 1999). In theory, a_i parameter, which is item discriminating index, has values around $(-\infty, +\infty)$. However in the ability test, negative discriminating items are excluded. For this reason it is stated that since a_i parameter doesn't get value more than +2, this parameter is valued around (0, +2) in applications (Hambelton and Swaminathan, 1985).

1.2. Problem Situation

As human characteristics are made subject in the measurements of education and psychology, development of measurement tools that are used in these fields reserve an important place in psychometrics. The purpose of the use of the techniques in the improvisation of measurement tools is to select the proper items and to produce quality tests as wanted (Baykul, 1991). As the test is being developed after written works are done in order to predict item characteristics, we try to get information about the structure of items by applying try-out practices. After having item analysis test is formed by calculating item statistics determining the items that are going to be placed in the test. In other words, in order to develop a test, which is as qualified as wanted, it is necessary to have try-out practices during the item analysis. However, there are situations, which may have objectionable features to apply try-out practices. Baykul (1991) stated that on the conditions the test should be confidential, it is often possible to come across with the situations where the entire test or the part of it is hidden. When secrecy cannot be provided or when there are worries about it, try-out practices are not applied. For instance, exams that are applied in Turkey such as the university entrance examination and high school entrance examinations that is organized by the Ministry of Education are not subject to try-out practices as the confidentiality has to be maintained. According to Sezer (1992), when try-out practices cannot be applied, if item statistics can be estimated with a fine approximation, tests with required characteristics can be obtained up to a point. As Baykul (1991) stated, on the conditions try-out practices cannot be applied item difficulty and item discriminating indexes can be estimated by the people who have knowledge about the group. By calculating the errors of these predictions, item difficulty indexes and item discriminating indexes can be estimated more accurately and the predicted statistics can be used as try-out results.

In this research, by estimating item characteristics based on experts' predictions when try-out practices cannot be applied, how the prediction power of item characteristics that is calculated based on classical test theory and two-parameter logistic model of latent trait theory will be determined. When try-out practices cannot be applied, studies on test development based on classical test theory that investigates how and in which ways test statistics can be predicted by estimating item statistics have been identified. However, today, studies based on latent trait theory, which gains growing importance in test developments, and studies in the comparison of the two theories have not been identified in literature. For this reason it is believed that this study will have a significant contribution to test development techniques in which try-out practices are not applied.

1.3. Problem Statement

What is the power of estimating the item characteristics based on the experts' predictions that are calculated with the aid of classical test theory and the two-parameter logistic model of latent trait theory when try-out practices cannot be applied?

1.4. Subproblems

1-What is the power of estimating item difficulty indexes-based on the classical test theory (pj)- by using the experts' predictions of the item difficulty indexes that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions?

2- What is the power of estimating the bi parameter- based on the two-parameter logistic model of the latent trait theory- by using the experts' predictions of the bi parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions?

3- What is the power of estimating the item discriminating power (rjx)- based on the classical test theory- by using experts' predictions of the item discriminating power indexes that belong to the mathematic sub-test of the student selection and placement examination of secondary school institutions?

4- What is the power of estimating the ai parameters -based on the two-parameter logistic model of the latent trait theory- by using the experts' prediction of the ai parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions?

2. METHOD

The aim of this study is to investigate the power of predicting item characteristics according to experts' predictions that are calculated based on classical test theory and the two-parameter logistic model of latent trait theory when the try-out practice cannot be applied.

2.1. Research Group

The research is applied to 9914 randomly selected students who have been placed in the 1999-2000 student selection and placement examination of secondary school institutions assigned by the Ministry of Education in Turkey.

2.2. Data Collection

The data of the research involves the answers to the mathematic sub-test of the students who took the student selection and placement examinations prepared by the Ministry of National Education's 1999-2000 education period for the secondary school institutions.

2.3.Data Analysis

In research, the effect of evaluation of values by the experts on predicted values by the experts on predicted values is analyzed by the use of estimation method argued in regression analysis. In the determination of research values, the first step taken is to calculate item parameter values separately according to classical and latent trait theories in order to depict the predicted values. With the benefit of the answers given to mathematic sub-test by the students, according to the classical test theory, item parameter values i.e. the strength of the item and item discriminating power indexes (pj and rjx) are calculated with the use of ITEMAN for window 3.1 program.

In the study before examining item parameter values of the mathematic sub-test (bi and ai) based on two-parameters logistic model of latent trait theory, it is tried to test whether the main assumptions which are normal distribution, one-dimensionality and local independence of latent trait theory are provided or not

Following the explanations about the basic assumptions of latent trait theory, item parameter values, which are derived from the mathematic sub-test -that are item difficulty and item discriminating power indexes (bi and ai)- are calculated by referring to latent trait theory and calculated by using the Bilog 3.0 package program.

After item parameter values of mathematical sub-test are calculated separately based on classical test theory and latent trait theory, it is asked to experts' predictions to estimate each item's parameter value on metric scale that are given to them based on classical and latent trait theories separately. Two different forms that are prepared to get experts' predictions are given firstly to 6 experts who prepare the mathematical sub-test of student selection and placement exams and then as the number of experts was not found sufficient, 10 mathematics teacher who offered preparation courses are involved in the study as well. In order to determine item parameters based on expert predictions, it is asked to put marks on item discriminating power index values of each test item approximately, on 10cm ruler which was given by the researcher before the test was applied. On the form, which contained parameters that was prepared based on the classical test theory, p_j parameter, which referred to the item difficulty index, and r_{jx} parameter, which is the item discriminating power index, were explained comprehensively to the experts. In addition, on the form, which was prepared based on the latent trait theory, the item difficulty index (b_i parameter) and the item discriminating power index (a_i parameter) were explained. With the thought that these explanations may not be satisfactory, education program was prepared by the researcher so as to help experts in the prediction of item parameters. The four-week education program was assigned -on different day and time scales- to experts; a one-week education was given to 6 experts and three-week education was arranged for 10 experts. In the education program, parameters that are contents of classical and latent trait theories were explained by the researcher to the experts and individual studies are carried out with the experts on the sample test items of different years whose item parameters were determined in advance. After studying the sample items of different years, when the experts were believed to understand the subject, they were asked to mark the item difficulty and discriminating power index values of the 1999-2000 mathematical sub-test items on the metric scale. The prediction values obtained from the metric scales of experts were measured by using a 10cm. ruler and each expert's prediction values of item parameters for classical and latent trait theories were determined separately. The studies showed that the estimation error of each expert in a group was greater than the estimation error, which was calculated by getting the average of estimations in that group and the reliability of an expert's estimation was found to be lower than the total estimation reliability and the error variance was large as a result of it (Sezer, 1992). As a result of the above explanations, since the consideration of each expert's prediction is believed not to support the aim of the study, the average of the item parameter values of 16 experts is taken.

After identifying the expert prediction values and the calculated prediction values of mathematic sub-test based on two different theories, regression method is used to identify the extent of the prediction power to which item characteristics based on the calculations of classical and latent trait theories can be determined by using the prediction power of item characteristics based on experts' predictions.

In order to solve the sub-problems of the research, after obtaining experts' prediction values on item characteristics that belong to mathematic sub-test, to determine prediction power of the item characteristics, which is calculated according to classical test and latent trait theories, regression analysis is administered by using SPSS 10.0 package program for windows statistics.

3. FINDINGS AND INTERPRETATION

In the research, before explaining the findings of sub-problems, it is thought to be appropriate to determine the reliability of experts' prediction values.

In order to determine the reliability of experts' prediction values, analysis of variance was applied for item parameter values, which were determined by 16 experts separately, based on the requirements of classical test and latent trait theories and a significant difference is not found between item difficulty index values of classical test theory and prediction values based on b_i parameters of latent trait theory of 16 experts. However a considerable difference is identified between item discriminating power index of classical test theory and prediction values based on a_i parameters of latent trait theory ($p < .01$). In other words, prediction values of item difficulty index parameters of

classical test and latent trait theories a consistency is acknowledged among experts whereas differences are identified in the discriminating power index values of experts' predictions.

3.1. What Is The Power Of Estimating Item Difficulty Indexes-Based On The Classical Test Theory (Pj) By Using The Experts' Predictions Of The Item Difficulty Indexes That Are Derived From The Mathematic Sub-Test Of The Student Selection And Placement Examination Of Secondary School Institutions?

To answer this problem, the average of item difficulty index of 16 experts' prediction values, belonging to 25 items, based on the mathematical sub-test and the item difficulty index values which are derived by using classical test theory are calculated.

In order to determine the relation between the average of experts' prediction values and observed item difficulty indexes, the product of pearson momentum correlation coefficient is calculated and whether the relation is more meaningfully different than 0 is tested. The coefficient of the relation between the average of experts' prediction values and observed item difficulty indexes is found to be 0.60. When the relation is tested on a 0.01 significance level, it is defined to be meaningful with the obtained result.

Foundation of a meaningful relationship between the average of pj values and the observed item difficulty indexes puts forth that, with the aid of the regression analysis, a regression function between the average of experts' prediction values and observed values can be formulated. In the study, before the application of regression analysis, the condition of data is determined with a pre-investigation study (Draper and Smith, 1981). After the pre-investigation stage, the provision of normal distribution of the dependent variable verified that the regression function of observed values can be formulated from the average of experts' prediction values. The statistics based on regression analysis are given in Table 1.

Table 1: Regression Analysis Results Of Observed pj Values Can Be Formulated From The Average Of Experts' pj Prediction Values

Coefficients	Estimate	Standart Error	t	Model R	SHE	F
a	0,09	0,07	1,27	0,60	0,11	13,2*
b	0,55	0,15	3,63			

*p<.01

Based on the regression analysis results, to estimate the observed item difficulty index from the average of experts' prediction values, the following regression function is formulated:

$$\hat{P}_j = 0,09 + 0,55.X \quad (2)$$

The sign of b coefficient in the regression function having a positive value refers to the fact that the change of experts' prediction values and observed values occurs in the same direction. Moreover, this coefficient indicates that as the independent variable of the average of experts' prediction values increase by one unit, observed item difficulty index values rise by 0,55.

Correlation of determination of the function is calculated to be 0,37 by getting the power correlation. 0,37 of the item difficulty index of the observed values can be explained by the average of experts' prediction values. In the table, the foundation of significant F values is also an evidence of the formulated function being appropriate. This value indicates that, the contribution of the independent variables to the variance in the dependent variable is vital however, when coefficient of determination is also considered, there may be other independent variables that effect the dependent variable.

When the analysis results are taken into account it can be said that, the experts' predictions of the item difficulty indexes that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions are capable of estimating the item difficulty indexes derived from the classical test theory.

3.2. What Is The Power Of Estimating The Bi Parameter- Based On The Two-Parameter Logistic Model Of The Latent Trait Theory- By Using The Experts' Predictions Of The Bi Parameter That Are Derived From The Mathematic Sub-Test Of The Student Selection And Placement Examination Of Secondary School Institutions?

To answer this problem, the average of bi parameters of 16 experts' prediction values, belonging to 25 items, based on the mathematical sub-test and the bi parameters which are derived by using the two-parameter logistic model of latent trait theory are calculated..

The correlation coefficient of the relation between experts' prediction values and observed bi parameters is calculated to be 0,68. When the relation is tested on a 0.01 significance level, it can be accepted as being meaningful with the obtained result.

Foundation of a meaningful relationship between the average of bi parameters and the observed bi parameters points out that, a regression function of observed values can be formulated by using the average of experts' prediction values. In the research, before the application of regression analysis, the condition of data is determined through a pre-investigation study.

By using the average of experts' prediction values, a regression function of the observed bi parameters is formulated and the findings about the statistics of regression analysis is given in Table2.

Tablo 2:Regression Analysis Results Of Observed Bg Values Can Be Formulated From The Average Of Experts' Bg Prediction Values

Coefficients	Estimate	Standart Error	t	Model R	SHE	F
a	0,88	0,34	2,60	0,68	1,50	19,6*
b	2,11	0,48	4,42			

*p<.01

As a result of regression analysis, in order to estimate the observed values of item parameter indexes from the average of experts' prediction values the following regression function is written.

$$\hat{b}_i = 0,88 + 2,11X \quad (3)$$

As the sign of regression coefficient is positive in the equation , it shows that there is a direct relationship between experts' prediction values and the predicted values. Also this coefficient proves that, as the average of experts' prediction value increases by one unit, bi parameter values increase by 2,11.

The coefficient of determination of the equation is found 0,46. As the coefficient of determination gives information about how much of the variance of dependent variable can explain independent variable, 0,46 of the variance in bi parameter values that is calculated according to two-parameter logistic model of latent trait theory can be explained by the average of experts' prediction values. As the F values in the table is found significant, it shows that the function is appropriate. This value proves that the contribution of independent variable to the variance of the dependent variable is essential.

When results of analysis is taken into account ,it can be said that bi parameter predictions based on experts' predictions that belong to mathematic sub-test of student selection and placement examinations have the capability of predicting bi parameters based on two-parameter logistic model of latent trait theory.

3.3. What Is The Power Of Estimating The Item Discriminating Power (R_{ij})- Based On The Classical Test Theory- By Using Experts' Predictions Of The Item Discriminating Power Indexes That Belong To The Mathematic Sub-Test Of The Student Selection And Placement Examination Of Secondary School Institutions?

To answer this problem, the average of r_{ij} parameter prediction of 16 experts', belonging to 25 items, based on the mathematic sub-test and the r_{ij} parameter values, which are derived by using classical test theory, are calculated.

In order to determine the relation between the average r_{ij} values based on experts' prediction and predicted r_{ij} parameters, correlation coefficient is calculated and found 0,16. When the relation is tested on a 0,01 significance level, it is defined not to be meaningful.

As a significant relationship between the average r_{ij} values based on experts and predicted r_{ij} parameters is not found prediction by using regression analysis is not done. According to findings that are obtained, it can be said that, r_{ij} parameter predictions depending on the prediction of the experts based on student selection and placement examinations do not have the capability of predicting r_{ij} parameters based on classical test theory. As mentioned in the introduction of the Findings' section, foundation of a significant difference between the item discriminating power index prediction values of experts' predictions i.e. inconsistency between their prediction values, may have affected the result of this finding. As, consistency between the item discriminating power index could not be obtained; relationship between the dependent and independent variables is negatively affected and meaningful estimations could not be obtained. The program that is applied by Quek (1989) in order to improve teachers' estimation ability is found effective on predicting item difficulty index, however it proves to be not effective on prediction of item discriminating power. Also according to Sezer's (1992) studies it is stated that it is more difficult to predict item discriminating power index comparison to item difficulty index. As well as in other studies it is determined in this study that even it is tried out on a group that is educated it is difficult to predict item discriminating power index.

3.4. What is the power of estimating the ai parameters -based on the two-parameter logistic model of the latent trait theory- by using the experts' prediction of the ai parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions?

To answer this problem, the average of ai parameter prediction of 16 experts', belonging to 25 items, based on the mathematic sub-test and the ai parameter values, which are calculated by using two-parameter logistic model of latent trait theory, are calculated. In order to determine the relationship between the average of experts' prediction values and predicted ai parameters the product of Pearson momentum correlation is calculated and found -0,22 when the relation is tested on a 0,01 significance level, it is defined not to be meaningful.

As the relationship between ai parameter prediction based on experts and the ai parameters, based on two-parameter logistic model of latent trait theory was meaningful and inverse directed it is proved that the predictions of the ai depending on experts that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions do not have the capability of predicting ai parameters, depending on two-parameter logistic model of latent trait theory. As, consistency between the ai prediction values could not be obtained; relationship between the dependent and independent variables is negatively affected and meaningful estimations could not be obtained. As it has been explained at the beginning of the findings section, it can be said that finding of connotational difference on ai assumption values among the experts, meaning that, inability of finding consistency between the experts' assumption values might have effected the finding result, as well. Not ensuring of consistency among the ai assumption values has negatively effected the relation between dependent and independent variable and has revealed that their estimations are pointless to predict. Although the findings that are obtained from the studies determined that the predictions of item discriminating power index based on classical test theory is difficult these findings proves that it is difficult to predict ai parameter that is defined as item discriminating power index in latent trait theory as well.

4. RESULTS

In the study, What is the prediction power of item characteristics calculated based on two parameter logistic model of latent trait theory is tried to be determined by estimating item characteristics based on experts predictions.

In the study before examining item parameters based on two–parameters logistic model of latent trait theory, it is tried to test whether the main assumptions of latent trait theory are provided or not.

In the first sub-problem of the study in order to determine what is the power of estimating item difficulty indexes-based on the classical test theory (p_j)-by using the experts' predictions of the item difficulty indexes that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions. According to findings, it can be said that, experts' predictions item difficulty indexes, which belong to mathematic sub-test of the student selection and placement examination of secondary school institutions, have the capability of predicting the item difficulty index values based classical test theory. As it is stated in this study, the experts who are involved in the research can have reliable predictions as they have a proper education on item difficulty index. Conditioned by the fact that content of the exam does not encounter any change, item difficulty index predictions can well be achieved.

The second sub-problem of the study aimed to explain whether by using the experts' predictions of the bi parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions, the prediction power of the bi parameter based on the two-parameter logistic model of the latent trait theory can be estimated.

In consideration with the regression analysis results, experts' predictions of the bi parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions, the prediction power of the bi parameter based on the two-parameter logistic model of the latent trait theory can be estimated. As in the study, after being involved in a thorough study, experts could provide a reliable estimation. Conditioned by the fact that, content of the exam does not change in the coming years, estimations of item difficulty index of latent trait theory which is known as bi parameter, can be attained.

The purpose of asking the third question is to explain if the item discriminating power (r_{jx}) based on the classical test theory can be estimated by using experts' predictions of the item discriminating power indexes that belong to the mathematic sub-test of the student selection and placement examination of secondary school institutions.

To solve this problem, first, the relationship between the average of experts' prediction values and observed r_{jx} parameters is was identified and correlation coefficient was calculated to be 0.16. When tested on a 0.01 significance level, the relation is concluded not to be meaningful. Thus, regression analysis was not applied in estimation. When the findings are checked, it can be concluded that item discriminating power (r_{jx}) based on the classical test theory cannot be estimated by using experts' predictions of the item discriminating power indexes that are derived from mathematic sub-test of the student selection and placement examination of secondary school institutions. In the other studies of the same field (Quek, 1989; Sezer, 1992), it is also found that although worked on an educated group, it is difficult to estimate the indexes of item discriminating power.

In the fourth sub problem of the study, by using the experts' prediction of the ai parameter that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions, what is the power of estimating the ai parameters -based on the two-parameter logistic model of the latent trait theory- tried to be determined. In order to identify the relationship between the average of experts' prediction values and the predicted ai parameters correlation coefficient is calculated and found $-0,22$. And when it tested at the 0.01 significance level, this relation is found meaningful. As the relationship was not found inverse directed it is proved that ai predictions, based on experts' that are derived from the mathematic sub-test of the student selection and placement examination of secondary school institutions, does not have the capability of predicting ai

parameters depending on two-parameter logistic model of latent trait theory. As, consistency between the ai prediction values could not be obtained; relationship between the dependent and independent variables is negatively affected and meaningful estimations could not be obtained.

REFERENCES

- Baker, B. F.(1977). Advances in item analysis. Review Of Educational Research, 47, 151-178.
- Baykul, Y. (2000). Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması, Ankara, Ösym Yayınları.
- Baykul, Y.(1991). “Ön denemenin yapılamadığı hallerde madde güçlük indeksinin ve buna bağlı olan madde ve test istatistiklerinin kestirilmesi.” Malatya:Üniversite Ve Çevre İlişkileri Sempozyumu.
- Draper, N.R. & Smith,H. (1981). Applied regression analysis, Wiley, Newyork.
- Edwards, L.A. (1984). An Introduction to linear regression and correlation, W.H.Freeman & Co: 2nd edition.
- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item-person statistics. Educational And Psychological Measurement, 58(3), 357-381.
- Hambelton, R.K & Swaminathan, H. (1985). Item response theory: Principles and application, Kluwer, Nijhoff Publishing A Member Of The Kluwer Academic Publisher Group,
- Hambleton, R.K; H. Swaminathan &H.J. Rogers.(1991). Fundamentals of item response theory, London, Sage Publications.
- Harvey, Robert J. & Allen L. Hammer.(1999). Item response theory. Counseling Psychologist, 27 (3), 353-374.
- Lord, F.M Ve M.R. Novick. (1968). Statistical theories of mental test scores, Addison Wesley Publishing Company, Educational Testing Service.
- Mislevy, J.R. & Bock, R.D. (1990). Bilog 3item analysis and test scoring with binary logistic models, Scientific Software.Inc.
- Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavı test kitapçığı. (2000). Milli Eğitim Bakanlığı Yayınları,.
- Quershı, M.Y & Fisher, Thomas L.(1977). Logical versus emprical estimates of item diffuculty. Educational And Psychological Measurement. 37, 91-100.
- Sezer, S.(1992). “Ön deneme yapılamayan durumlarda madde güçlük ve ayırıcılık gücü indekslerinin ve bunlara bağlı test istatistiklerinin kestirilmesi “ (Yayınlanmamış Doktora Tezi), Ankara: H.Ü. Sosyal Bilimler Enstitüsü.
- Stocking, M.L. (1997). Item response theory. Educational Research, Methodology And Measurement: An International Handbook. Pp.836-839.Edited By: John P.Keeves .Second Edition. Elsevier Science Ltd.
- Stevens, S. S. (1946). On the theory of scales of measurement. Science,103, 677-680.
- Thorndike, R. L. (1971). Educational measurement, Second Edition.
- Torgerson, W. S. (1958). Theory and methods of scaling. Newyork: John Willey.
- Turgut, M. F. (1997). Eğitimde ölçme ve değerlendirme metodları, Ankara, Yargıcı Matbaası.
- Turgut, M. F. ve Baykul, Y. (1992). Ölçekleme teknikleri,Ankara,ÖSYM Yayınları.
- Weitzenhoffer, A.M.(1951). Mathematical structures and psychological measurement. Psychometrika, 16, 387-406.

GENİŞLETİLMİŞ ÖZET

Bu araştırmada, deneme uygulaması yapılamayan durumlarda madde özelliklerinin uzman tahminlerine dayalı kestirimlerinin klasik test teorisi ve örtük özellikler teorisinin iki parametrelili logistik modeline dayalı olarak hesaplanan madde özelliklerini tahmin etme gücünün nasıl olduğu incelenmeye çalışılmıştır. Araştırma, Türkiye’de uygulanan Milli Eğitim Bakanlığının 1999-2000 öğretim yılı Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavına giren random olarak seçilmiş 9914 öğrenci üzerinde yapılmıştır. Çalışmada uzman tahmin değerleri; 16 matematik öğretmenin klasik ve örtük özellikler test teorilerine göre ayrı ayrı hazırlanan iki metrik ölçek formuna verdikleri kestirim değerlerinden elde edilmiştir.

Araştırmada, uzmanların maddelere ilişkin tahmin değerlerinin gözlenen değerler üzerindeki etkileri ön kestirim yöntemi olan regresyon tekniğinden yararlanılarak analiz edilmiştir. Araştırmanın verilerinin çözümlenmesinde öncelikle gözlenen değerleri belirlemek için klasik ve örtük özellikler teorilerine göre ayrı ayrı madde parametre değerleri hesaplanmıştır. Öğrencilerin, Matematik alt testine ilişkin vermiş oldukları cevaplardan yararlanarak klasik test teorisine göre madde parametre değerleri yani madde güçlüğü ve madde ayırıcılık gücü (p_j ve r_{jx}) indeksleri hesaplanmıştır. Örtük özellikler teorisine göre Matematik alt testinin madde parametre değerleri (b_g ve a_i) hesaplanmadan önce, bu teorisinin temel varsayımları olan normal dağılım, tek boyutluluk ve yerel bağımsızlık varsayımlarının karşılanıp karşılanmadığı test edilmiştir.

Araştırmanın ilk alt probleminde, Türkiye’de uygulanan Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait madde güçlük indekslerinin uzman tahminlerine dayalı kestirimlerinin klasik test teorisine dayalı madde güçlük indekslerini (p_j) tahmin etme gücünün nasıl olduğunu belirlemeye çalışılmıştır. Elde edilen bulgulara göre, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait madde güçlük indekslerinin uzman tahminlerine dayalı kestirimlerinin klasik test teorisine dayalı madde güçlük indeks değerlerini yordayabildiği söylenebilir. Bu çalışmada olduğu gibi, tahminde bulunacak uzmanlar madde güçlük indeksi hakkında yeterli bir eğitimden geçirildiklerinde güvenilir bir şekilde tahminde bulunabilmektedirler. Böylelikle madde güçlük indeksinin tahmin değerlerinden yararlanılarak klasik test teorisine dayalı madde güçlük indeksleri kestirilebilmekte ve uzmanlara ait kestirim denklemi elde edilmektedir. Sınavın içeriğinde bundan sonraki yıllarda değişiklik olmaması şartıyla bulunan bu denklemden yararlanılarak madde güçlük indekslerinin tahminlerinin yapılabileceği bu çalışmayla ortaya konmuştur.

Araştırmanın ikinci alt probleminde; Türkiye’de uygulanan Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait uzman tahminlerine dayalı b_i parametre kestirimlerinin örtük özellikler test teorisinin iki parametreliliğe dayalı b_i parametrelerini tahmin etme gücü nasıl olduğu belirlenmeye çalışılmıştır. Analizi sonuçları dikkate alındığında, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait uzman tahminlerine dayalı b_i parametre kestirimlerinin örtük özellikler test teorisinin iki parametreliliğe dayalı b_i parametrelerini yordayabildiğinden söz edilmiştir. Bu çalışmada olduğu gibi, tahminde bulunacak uzmanlar b_i parametresi hakkında yeterli bir eğitimden geçirildiklerinde güvenilir bir şekilde tahminde bulunabilmektedirler. Elde edilen uzmanlara ait kestirim denklemlerinden yararlanılarak, sınavın içeriğinde bundan sonraki yıllarda değişiklik olmaması şartıyla örtük özellikler teorisinde madde güçlük indeksi olarak tanımlanan b_i parametresi tahminlerinin yapılabileceği bu çalışmayla ortaya konmuştur.

Araştırmanın üçüncü alt probleminde Türkiye’de uygulanan Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait madde ayırıcılık gücü indekslerinin uzman tahminlerine dayalı kestirimlerinin klasik test teorisine dayalı madde ayırıcılık gücünü (r_{jx}) tahmin etme gücü nasıl olduğu belirlenmeye çalışılmıştır. Analiz sonuçları incelendiğinde, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait uzman tahminlerine dayalı r_{jx} parametre kestirimlerinin klasik test teorisine dayalı r_{jx} parametrelerini yordayamadığından söz edilebilir. Bulgular bölümünün başında açıklandığı gibi uzmanlar arasında madde ayırıcılık gücü tahmin değerlerinde manidar fark bulunmuş olması yani uzmanların tahmin değerleri arasında tutarlılığın bulunmaması olmasının da bu bulgu sonucunu etkilediği düşünülmektedir. Madde ayırıcılık gücü indeksinin tahmin değerleri arasında bir tutarlık sağlanamaması bağımlı değişkenle bağımsız değişken arasındaki ilişkiyi olumsuz yönde etkilemiş ve kestirimlerinin yordamada bulunmak için anlamlı olmadığını ortaya koymuştur. Bu çalışmada olduğu gibi literatürde yapılan çalışmalarda (Quek,1989; Sezer,1992) eğitim verilmiş bir grup üzerinde çalışılmış olsa bile madde ayırıcılık gücü indeks kestirimlerinin zor olduğunu ortaya koymuştur.

Araştırmanın dördüncü alt probleminde, Türkiye’de uygulanan Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait madde ayırıcılık gücü indekslerinin uzman tahminlerine dayalı kestirimlerinin örtük özellikler test teorisinin iki parametreliliğe dayalı a_i parametresini tahmin etme gücü nasıl olduğu belirlenmeye çalışılmıştır. İlişkinin ters yönlü ve anlamlı bulunmamış olması, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testine ait uzman tahminlerine dayalı a_i kestirimlerinin örtük özellikler test teorisinin iki parametreliliğe dayalı a_i parametrelerini yordayamadığını ortaya koymuştur. Bulgular bölümünün başında açıklandığı gibi uzmanlar arasında a_i tahmin değerlerinde manidar fark bulunmuş olması yani uzmanların tahmin değerleri arasında tutarlılığın bulunmaması bu duruma yol açmış olabilir. a_i tahmin değerleri arasında bir tutarlık sağlanamaması bağımlı değişkenle bağımsız değişken arasındaki ilişkiyi olumsuz yönde etkilemiş ve kestirimlerinin yordamada bulunmak için anlamlı olmadığını ortaya koymuştur.