International Journal of Progressive Education, Volume 15 Number 4, 2019

An Investigation of Pre-Service Preschool Teachers' Projects Using The Many-Facet Rasch Model *

Durmuş Özbaşı [i]
Çanakkale Onsekiz Mart University

Serdar Arcagök [ii]
Çanakkale Onsekiz Mart University

Abstract

The aim of this study was to evaluate project proposals prepared by pre-service preschool teachers' using ten criteria and thus determine the awareness of pre-service preschool teachers' toward the Research Project course. The survey method, a quantitative research type, was used for this research. The study group of the research constituted six different project proposals prepared by final year undergraduate students taking the Research Project I and Research Project II courses at the Faculty of Education, Department of Pre-school Education of Çanakkale Onsekiz Mart University, Turkey during the academic year 2018-2019. Twelve academic members of the Faculty of Education from six different universities assessed the projects. The data collection tool used was based on the research project assessment criteria of TÜBİTAK (Scientific and Technological Research Council of Turkey) along with the course's learning outcomes and content, and the theoretical frame of the research topic. The results were analyzed based on the Many-Facet Rasch Model. According to the data obtained, it was found that the projects and project criteria differed in terms of consistency and generosity. It was also observed that the pre-service teachers met some of the criteria while they had difficulties with other criteria.

Keywords: Many-facet rasch analysis, pre-service preschool teachers, research projects, survey method.

DOI: 10.29329/ijpe.2019.203.12

----
*This paper was presented as a summary paper at the 12th International Congress of Educational Research, 25-28 April. 2019, Rize, Turkey

[i] **Durmuş Özbaşı**, Assist. Prof. Dr., Canakkale Onsekiz Mart University, Faculty of Education, Measurement and Evaluation in Education.

Correspondence: dozbasi@gmail.com

[ii] **Serdar Arcagök**, Assist. Prof. Dr., Çanakkale Onsekiz Mart University, Faculty of Education.

INTRODUCTION

The flexible development of individuals is of great importance in our era where information and communication technologies are rapidly changing (Akgün, 2000). Individuals are expected to be equipped with the qualifications required by this era, which has been described the Digital Era; in other words, individuals need to possess the most recent knowledge and skills. The skills expected from such qualified individuals include being able to sense the problems they face and express them, to be able to hypothesize depending on the nature of the problem, to detect possible variables, to propose possible relations between variables and to be able to define relationships clearly and explicitly (Yaşar, 2014). From this point of view, it is necessary that individuals gain science processing skills, which is one of the thinking skills. Science process skills can be defined as a process comprising various stages such as observing qualities, measuring quantities, sorting, classifying, inferring, predicting, experimenting, communicating, modeling, changing variables and controlling them (Turkish Ministry of Education - MEB, 2005). The aforementioned process can be conducted functionally in the case of individuals who have a scientific attitude. Those having a scientific attitude should conduct their research in a systematical, skeptical and ethical manner. Systematicity signifies that the individual has the tendency to approach issues seriously and question what, how and why. Being skeptical means knowing the possibility that the idea put forward may not be verified. Taking into consideration individuals who undertake research, and their concerns and rewards, acting in compliance with the possibility that they may be affected by following or observing certain behavioral rules, denotes being ethical (Robson, 2015).

It is also essential that a research approach covers philosophical assumptions in addition to different methods and process steps. In other words, those preparing a research plan or proposal should possess broad philosophical views. These views generally consist of elements of post positivism, constructivism, transformativism and pragmatism. Post-positivists adopt a causal philosophy in which reasons determine possible results or products. Constructivism describes an individual's search for meaning about the world they live in or the world they experience. The transformative philosophical assumption is an assumption defended by those who consider that the structural rules and theories imposed by post-positivists in the 1980s and 1990s do not correspond to individuals marginalized in society or issues such as power and social justice, discrimination or pressure, which need to be investigated. On the other hand, adherents of pragmatism generally take into consideration actions, situations and results rather than initial circumstances (Creswell, 2014).

In the Digital Era in which we live, undergraduates' ability to gain adequate knowledge, skills and attitude in conducting studies will enable them to occupy positions in different business fields and to use efficiently the outcomes they have acquired (İlhan, Çelik and Aslan, 2016). Studies conducted while they are in the educational phase make a crucial contribution to acquiring the necessary knowledge, skills and attitudes.

Undergraduates need to have research literacy due to the importance of conducting research in education and society. Undergraduates will be able to find articles about their field and will be able to evaluate those articles and then suggest and conduct research studies at any time in their own career; leading to an improvement in the awareness of undergraduates of educational research (Johnson and Christensen, 2014). On the other hand, research in education will encourage students to follow closely the changes and developments in their own professional fields and to become a qualified researcher.

Although there is no consensus on differentiating the kinds of educational research among the experts and scientists, research conducted in this field has mostly been divided into four types. These can be listed as follows (Best and Kahn, 2017):

1. Historical Research compromises the examination of past events or incidents to research, record, analyze and evaluate where, when, with whom or which bodies they had happened (Demirel, 2008; Ekiz, 2009).

2. Quantitative Descriptive Research tests the accuracy of theories and examines the relations within a certain structure. These studies are based on numbers and symbols and the research results can be generalized (Büyüköztürk, Çakmak, Akgün, Karadeniz and Demirel, 2009).

3. Qualitative Descriptive Research enables a deep and comprehensive investigation of a subject or events. The researcher themselves is usually the data collection tool (Patton, 2014).

4. Experimental Research aims to explain the circumstances causing an event and affecting that event; in which the relations between one or more variables are scrutinized (Ary, Jacobs, Razahiev and Sorensen, 2006).

Studies in the educational field comprise processes enabling teachers to learn how to ask questions based on the studies they conduct, how to organize the research method, how to analyze data and report the results, and how to cooperate with other researchers. Thus, teachers and preservice teachers conduct accurate and objective research and follow a scientific path (Herman, Clough and Olson, 2013; Schwarz, Westerlund, Garcia and Taylor, 2010). These studies provide an understanding of the functionality and extent of different shareholders such as the learners, teachers and administrators. In this sense, these studies aim to describe, predict, organize and explain the mentioned shareholders (Gall, Gall and Borg, 2007). In other words, studies in education aim to keep alive the constant questioning skill, to produce new ideas, to develop theories related to the field of application, and to define the productivity and efficiency of curriculums (Mahoney, 2013).

Studies conducted on education also make several contributions for preservice teachers in various contexts. These include obtaining and giving feedback to preservice teachers about how to best apply and manage curriculums, evaluating pedagogical field knowledge, and assessing the functionality of curriculums (Zientek, Capraro and Capraro, 2008). Studies enable teachers and preservice teachers to acquire ethical and epistemological values. In addition, they are also encouraged to adopt an objective perspective by helping them to acquire democratic values (Murray, 2017).

The functionality of studies conducted about education can be achieved by increasing the awareness of preservice teachers towards these studies. Therefore, it is essential to ensure that preservice teachers assess studies conducted in the field with a questioning point of view and gain research experience in the field (Creswell and Plano-Clark, 2010).

It is pointed out by the Ministry of Education (MEB) that one of the field knowledge competencies that a teacher should have regarding the teaching profession should cover the questioning perspective so as to incorporate methodological knowledge as well as theoretical and factual information. This has also been shown to be the indicator of fundamental competence to categorize basic methods and techniques (MEB, 2017). Correspondingly, the Council of Higher Education (YÖK, 2007) emphasizes that one of the qualifications for lecturers teaching undergraduate programs is to train preservice teachers to conduct scientific research and to make use of the findings. Therefore, the one-semester "Research Project I" and "Research Project II" courses have each been included within the undergraduate programs of pre-school teaching so as to enable preservice teachers to acquire a questioning perspective, to follow studies conducted in education, and to carry out their own research studies in the field (YÖK, 2007). Via the Research Project II course, it is expected that the preservice teacher can define a research topic related to the field of education, compose research questions relating to the research, adopt a methodology appropriate to the topic and report the data collected. In addition, it is expected that the importance of the studies conducted in the field of education is adopted by the teacher as well. Indeed, YÖK included the Project Preparation course as an elective course within the curriculum as professional knowledge for pre-school teachers in the 2018-2019 academic year by stressing the importance of studies in the field of education (YÖK, 2018).

Importance of Research

When the related literature was examined, it was found that studies conducted with undergraduate students in the field of education were mostly related to scientific research methods (Akar, 2007; Akkanat, Abu, Çakır and Gökdere, 2017; Aksu, 2018; Bins, 2009; Ersoy, 2016; Garza, 2015; Hypolite, 2012; İlhan, 2016; Orçan, 2013; Spang, 2008; Yaşar, 2014). Nevertheless, studies examining undergraduate students' scientific process skills can also be seen (Aydoğdu, 2009; Çelik, 2013; Kaya and Yılmaz, 2016; Kefi, Çeliköz and Erişen, 2013; Kılıç, Haymana and Bozyılmaz, 2008). However, it was seen that the number of studies conducted with undergraduate students on the Research Project course was limited (Cengiz and Karataş, 2014; Eti and Gündoğdu, 2015) and that no study had been conducted using the Many-Facet Rasch model. Curriculums in Turkey have been based on the constructivist approach since 2005. The program, having a student-centered and a helical structure, supports the versatile development of students. One of the key components of this approach is supporting the creative and critical thinking and science process skills of students. This can only be achieved given that the preservice teachers whose aim is to train the next generation also acquire these skills. Therefore, the Research Project course considerably contributes to preservice teachers producing authentic ideas with a critical perspective. Besides this, preservice teachers receive knowledge and gain awareness regarding scientific research methods and develop their science process skills. Additionally, this course helps preservice teachers to learn the fundamental stages of research proposals by applying them. It is therefore believed that measuring the efficiency of this course using Many-Facet Rasch analysis will make a considerable contribution to the related literature.

Theoretical Framework

Item Response Theory and Many-Facet Rasch Model

The Item Response Theory (IRT) is a mathematical model proposed and developed as a reaction to the Classical Test Theory (CTT) to minimize its weaknesses (Hambleton, Swaminathon and Rogers, 1991). One of the most important advantages of IRT over the CTT is that IRT is able to make predictions by eliminating individual and group influence within the frame of the invariance principle (Hambleon, 1995). Consequently, according to IRT, while making predictions related to item difficulty and item differentiates, which are two psychometric qualities, it does not matter in which group the study is conducted. The IRT has four different models, namely, one-, two-, three- and four-parameter. Only the one-parameter logistic model constituting the basis of the Many-Facet Rasch model will be mentioned here as it is the most used within the scope of this study.

The one-parameter logistic model (1PLM) is a model that only covers one item of difficulty parameter. In this model, the item differentiate indexes of all items are considered to be equal. The item characteristic curves are also the same for all items in 1PLM. The one-parameter model is referred to by the name of its developer, George Rasch. Linacre (1989) developed the Rasch model by reducing the rater effect. This model compromises several variables such as the rater, scoring, items/features and is also known as the Many-Facet Rasch Model (MFRM) in the literature (Mulqueen, Baker and Dismukes, 2000; Chapman, Letourneau and Sheidow, 2013). MFRM is shown by the formula below:

$$\text{Log}\,(P_{nijk}/P_{nijk-1}) = \theta_n - D_i - C_j - F_k$$

P_{nijk} : Probability of examinee n receiving a rating of k on criterion i from rater j

θ_n : Proficiency of examinee n

D_i : Difficulty of criterion i

C_j : Severity of rater j

F_k : Difficulty of receiving a rating of k relative to a rating of k – 1

MFRM is a model including all sources of variability that are thought to influence the scores in the analysis and showing the interaction between these sources of variability (Kim, Park and Kang, 2012). According to this, there are many sources of variability or facets, such as “examinee x item”, “item x rater”, and “rater x examinee”, etc. As the MFRM is an extension of IRT, the item difficulty, examinee’s scores and assumptions regarding raters can be conducted independently of the group or separately, and then these can be degraded to a common criterion level (with the data calibration map) and all facets can be interpreted simultaneously (O’Neil and Lunz, 1996; Kim, Park and Kang, 2012). This aspect provides a great advantage for researchers. More specifically, the MFRM is a model that enables the comparison of potential interactions between facets, the severity/leniency of raters, the degree of rater consistency, and item difficulty levels by bringing examinees, skills, items and raters to a common measurement level (Sudweeks, Reeve and Bradshaw, 2004; Güler, 2008; Yue, 2011; Linacre, 2014)

Aim

The general aim of the study was to have the project proposals prepared by students assessed with 10 criteria by the judges and to examine the judges, criteria and project facets by means of the Many-Facet Rasch Model. Answers to the following questions were sought within the scope of the general aim.

1. What is the condition of the calibration map obtained for “rater, project and criterion” facets in the scoring carried out, as per the Project Evaluation Criteria?

2. What are the statistics regarding the measurement report of the project proposals?

3. What are the statistics of the measurement report of the criteria used in assessment of the project proposals?

4. How does the consistency/severity of the judge change during the scoring at the assessment stage of the project proposals?

5. Are there any biased interecation between projects and judges’ in the scoring?

METHOD

Research Model

The survey model was used in the study. This model aims to collect data from a wide sample during a certain period (Best, 1998). Another aim of the survey model is to analyze a current situation by defining and explaining it (Ekiz, 2009).

Study Group

The study group of the research constituted six different project proposals prepared by final year pre-service preschool teachers taking the Research Project I and Research Project II courses at the Faculty of Education, Department of Preschool Education of Çanakkale Onsekiz Mart University, Turkey during the academic year 2018-2019. The assessors taking part in the study were 12 faculty members working at six different universities in Turkey, namely, Fırat University, Trakya University, Samsun Ondokuz Mayıs University, Süleyman Demirel University, Kilis 7 Aralık University and Çanakkale Onsekiz Mart University.

Data Collection Tool

The data collection tool used in the research was developed in consideration of TÜBİTAK's research project evaluation criteria along with the course's learning outcomes and content, and the theoretical frame of the research topic. The research project evaluation criteria (data collection tool) created were submitted for the approval of five academic members teaching the Research Project course as part of the undergraduate program and having realized a research project with their undergraduate students. Upon the evaluation of the five academic members by calculating the Content Validity Index (CVI), it was decided whether or not to use each item in the project proposal survey tool as criteria. The CVI was used in order to determine whether there was coherence among the experts (Yurdugül 2005; Lawshe, 1975). The CVI was calculated using the formula below:

$$KGO=\frac{UG}{N/2}-1$$

UG: number of experts sharing the "appropriate" view of the item

N: Total number of experts

The five academic members were asked to assess the evaluation criteria used within the scope of the study as "appropriate", "needs correction" or "not appropriate" regarding whether the mentioned criteria could be used as project proposal evaluation criteria or not. According to this, the CVI of the 10 criteria chosen in compliance with the views of the five academic members was calculated as 1. Consequently, the ten project proposal evaluation criteria were defined as: creating a project title, writing a project abstract, determining key words found in the project abstract, forming the theoretical frame of the project, revealing the authentic value of the project, defining an event calendar for the project, indicating the common effect expected from the project, creating a general budget for the project, and justifying the project budget.

Analysis of Data

Analysis of the data collected regarding the student projects was conducted within the frame of MFRM and realized with the FACETS program developed by Linacre (2014). MFRM is an extension of the Rasch model based on the IRT. The study was designed as a three-facet model compromising the judge, evaluation criteria and projects. According to this, the data calibration map and the three facets were evaluated with the same criteria. In addition, the scoring of the judge, criteria and projects was calculated to produce detailed findings on each facet. Detailed information about the judge and project interaction was also obtained by analyzing the severity and consistency of the judges.

It is also essential to test the assumptions before conducting analysis based on the Many-Facet Rasch Model, as the Rasch model is based on the IRT (Baker, 2001). These assumptions are (a) unidimensionality, (b) local independence, and (c) data-model fit.

a) Unidimensionality

The assumption of unidimensionality should be tested to perform the Many-Facet Rasch Model and to properly interpret the findings obtained. Unidimensionality can be described as measuring the target psychological feature under a mode factor (Hambleton, Swaminathon and Rogers, 1991). Exploratory Factor Analysis (EFA) was used to define whether the survey tool was unidimensional or not. EFA is an analysis technique used to define the latent sources of the variance and co-variance in the data obtained and to explain them (Jöreskop and Sörbom, 1993). The Kaiser Mayer Olkin value regarding the adequacy of the sample was found to be .63 and the Bartlett globality test also found it to be statistically meaningful ($\chi 2$ (45)=129,611; p<.01). Accordingly, it can be said that the data is in compliance with the factor analysis. The obtained EFA results are shown in Table 1.

Table 1 EFA Results For Project Evaluation Tool

Criterion no.	Factor Load	Criterion no.	Factor Load	Criterion no.	Factor Load
CRT1	.30	CRT5	.31	CRT9	.30
CRT2	.71	CRT6	.66	CRT10	.58
CRT3	.52	CRT7	.30		
CRT4	.57	CRT8	.69		
Attribute = 2.546, Announced Variance = 30%					

As per the EFA results given in Table 1.1, the criteria in the project evaluation tool explain 30% of the total variance under a single factor. The factor load values vary from .71 to .30. In this sense, it can be said that the project evaluation tool features unidimensionality.

b) Local Independence

Local independence indicates whether there is a relationship between the response to a survey tool and the response to another item, and is frequently associated with unidimensionality (Hambleton, Swaminathan and Rogers, 1991; de Ayala, 2009). The fact that the test meets the unidimensionality assumption is sufficient to assume local independence. Thus, in our study the unidimensionality assumption was met and no additional analysis for local independence was required.

c) Data-Model Fit

The data-model fit is related to how small the standardized residual value (StRes) is. It is also called the "unexpected value". According to this, for values outside a range of ±2 of the standardized residual value, it should not be more than 5% of the entire data to determine the data-model fit. This value should not be over 1% of the total data for values in the range of ±3 of the standardized residual value (Linacre, 2003). In the present study, the standardized residual values were examined to check whether the data-model fit was met. The residual ratio for values outside the range of ±2 was 2.9%, while the standardized residual ratio for values outside the range of ±3 was 2.9%. Consequently, it was concluded that the data-model fit was appropriate and the analysis could be continued.

RESULTS

The findings obtained in the study are discussed as per the sub-aims of the study.

Findings related to first sub-aim

The first sub-aim of the study was "What is the condition of the calibration map obtained for "rater, project and criterion" facets in the scoring carried out, as per the Project Evaluation Criteria?" Accordingly, 6 project reports were assessed by 12 different raters in terms of 10 criteria and scored. The scoring results were analyzed according to the Many-Facet Rasch Model.

As per the iteration results of the analysis, it was found that 18 iterations were conducted. The low number of iterations shows that it is easy to obtain a good prediction from the data (İlhan, 2015). A variable (calibration) map of the facet statistics based on the study's data was formulated and is shown in Figure 1.

```
+------------------------------------------------------------------------+
|Measr|+project|-crt          |+project|-judge                     |PROJE|
|-----+--------+--------------+--------+---------------------------+-----|
|   2 +        +              +        +                           + (4) |
|     |        |              |        |                           |     |
|     |        | crt8         |        |                           |     |
|     |        |              |        |                           | --- |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|   1 +        + crt5         +        +                           +     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |  3  |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     | 5      | crt4         | *      |                           |     |
|     |        | crt2         |        |                           |     |
|     | 2 4    |              | **     |                           |     |
*   0 * 6      * crt7         * *      *                           *     *
|     |        |              |        |                           | --- |
|     | 1 3    | crt6         | **     | judge3   judge4           |     |
|     |        | crt1    crt10|        | judge5                    |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        | judge1                    |     |
|     |        |              |        |                           |  2  |
|     |        |              |        | judge2   judge6   judge8  |     |
|  -1 +        +              +        + judge10  judge9           +     |
|     |        | crt9         |        | judge12  judge7           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        | crt3         |        |                           |     |
|     |        |              |        | judge11                   |     |
|     |        |              |        |                           | --- |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|     |        |              |        |                           |     |
|  -2 +        +              +        +                           + (1) |
|-----+--------+--------------+--------+---------------------------+-----|
|Measr|+project|-crt          | * = 1  |-judge                     |PROJE|
+------------------------------------------------------------------------+
```

Figure 1. Data Calibration (variable) Map

When Figure 1 is examined, it can be said that the project with the highest proficiency level is Project 5 (0.30 logit), and the projects with the lowest proficiency levels are 1 and 3 (-0.20 logit, respectively). According to Figure 1, the criteria used in the project evaluation are easier from the top to the bottom and more difficult from the bottom to the top. Consequently, the hardest criterion for students while they were writing their projects was criteria 8 (1.8 logit). On the judge facet of the data calibration map, the judge located at the top was the most severe while the judge at the bottom was less severe in terms of scoring. According to the judge column in Figure 1.1, it can be said that the most severe judges were 3 and 4 (-0.2 logit). The most generous judge was judge 11 (logit = -1.5). It is necessary to examine the survey reports pertaining to the project, criterion and judge facets in order to examine each facet in more detail. Table 2 shows the results regarding the students' projects.

Table 2 Student projects measurement report

Project	Observed Average	Fair Average	Model Measure	Model Error	Infit Square Average	Infit Z	Outfit Square Average	Outfit Z
P5	3,17	3,26	.29	.13	1,16	-.3	1,13	.9
P2	3,07	3,15	.07	.13	,95	-.6	,95	-.3
P4	3,07	3,15	.07	.13	,91	.3	,95	-.3
P6	3,02	3,10	-.02	.13	1,04	-.2	1,13	.9
P3	2,92	3,01	-.19	.12	,96	-.5	1,00	.0
P1	2,91	2,99	-.22	.12	,93	-.3	,91	-.6
Average	3,03	3,11	.00	.13	,99	,0	1,01	,1
Standard Deviation	,09	,09	.19	.00	,09	,6	,1	,6

Model, Sample: RMSE .13 Adj (True) S.D. .14 Separation 1.14 Strata 1.85 Reliability .70
Model, Fixed (all same) chi-square: 11.3 d.f.: 5 significance (probability): .05

Table 2 gives details of the Many-Facet Rasch analysis results regarding the facets, including the students' project evaluations. When the RMSE values of the logit values were examined, it was calculated as 0.13. On the other hand, the standard deviation of the RMSE value is below 1 (0.14), which is a critical value. The relevance coefficient was calculated as 1.14, and the reliability index was calculated as 0.70. The reliability index regarding the facets in the Many-Facet Rasch analysis is not interpreted as in the CTT. The high value of the reliability index regarding facets in the Many-Facet Rasch analysis gives information about the difference between the raters in terms of severity/leniency (Haiyang, 2010). According to this, the fact that the reliability index is high does not mean that the raters display similarities but it indicates that the difference is reliable. In other words, the fact that this difference is large means that the projects are similar or different in terms of similarity and difference. If the reliability index is high, it indicates that the students' projects are similar; whereas if the reliability index is low, the students' projects are different. When the chi-square value obtained in the study is examined, it can be said that the difference is meaningful ($\chi 2=11.3$, sd=5, $p<.05$). Therefore, the absence hypothesis was rejected in the hypothesis that "there is a meaningful difference in students' projects" about the constant effect. In that case, it can be concluded that there is a statistically meaningful difference between the rater judges and the students' projects.

When the infit and outfit statistics of the facets are examined, it can be concluded that the reference range of all 6 projects was from 0.6 to 1.4 (Wright and Linacre, 1994: 375-380) and there is no facet exceeding those values.

Rater Analysis

It is necessary to examine separately each facet used in the project regarding the logit values for the criteria and judge facets in order to assess each facet in more detail. Thus, the facet statistics on the judges' evaluation, as the raters in the study, are given in Table 3.

Table 3 The Measurement report of judges

Judge	Observed Average	Fair Average	Model		Infit		Outfit	
			Measure	Error	Square Average	Z	Square Average	Z
judge4	2,68	2,73	-0,18	0,17	0,99	0	1,08	0,5
judge3	2,72	2,77	-0,24	0,17	0,52	-3,3	0,55	-3
judge5	2,75	2,81	-0,29	0,17	0,81	-1	0,9	-0,5
judge1	2,98	3,06	-0,7	0,18	0,63	-2,3	0,64	-2,2
judge6	3,07	3,14	-0,86	0,18	2,08	4,5	1,88	3,8
judge8	3,07	3,14	-0,86	0,18	0,87	-0,6	0,91	-0,4
judge2	3,08	3,16	-0,89	0,18	1,16	0,8	1,37	1,8
judge9	3,12	3,2	-0,96	0,18	1,1	0,5	1,05	0,3
judge10	3,15	3,23	-1,02	0,18	1,11	0,6	1,01	0,1
judge12	3,17	3,25	-1,06	0,19	0,96	-0,1	0,89	-0,5
judge7	3,18	3,26	-1,09	0,19	0,75	-1,3	0,69	-1,7
judge11	3,35	3,43	-1,47	0,2	1,04	0,2	1,16	0,8
Average	3,03	3,1	-0,8	0,18	1	-0,2	1,01	-0,1
Standard Deviation	0,2	0,21	0,37	0,01	0,38	1,9	0,34	1,7

Model, Sample: RMSE .18 Adj (True) S.D. .34 Separation 1.91 Strata 2.88 Reliability (not inter-rater) .78
Model, Fixed (all same) chi-square: 52.5 d.f.: 11 significance (probability): .00
Model, Random (normal) chi-square: 9.1 d.f.: 10 significance (probability): .52

The logit, infit, outfit values and reliability index of the rater facet are illustrated in Table 3.When the RMSE value indicating the standard error of the logit values in the table was examined, it was calculated as 0.18 and the standard deviationdevaition of the RMS value was calculated as under 1.00 (0.34), which is the critical value. The relevance coefficient was calculated as 1.80 and the reliability index as 0.76. The calculated reliability index gives information about the difference between raters in terms of severity/leniency (Haiyang, 2010). According to this, the fact that the reliability index is high does not mean that the raters feature similarities, but that the difference is reliable. It is essential to examine the chi-square results to conclude whether the difference is meaningful or not. When the chi-square value obtained in the study is examined, it can be said that the difference is meaningful ($\chi2=52.5$, sd=11, p<.01). In other words, the absence hypothesis was rejected in the hypothesis that “there is a meaningful difference in students’ projects” about the constant effect. In this case, it can be concluded that there is a statistically meaningful difference between the judge raters and the students’ projects.

In addition, according to the facet analysis regarding judges given in Table 3, when the infit and outfit values of the 12 judges are examined, it was concluded that 11 of the judges’ values were within the reference range proposed by Wright and Linacre (1994: 375-380), being 0.6 to 1.4, and only judge 1 was outside this reference value.

Analysis of criteria used in project evaluation

The survey findings regarding the Many-Facet Rasch analysis on the compliance of criteria used by judges to evaluate students’ projects are given in detail in Table 4.

Table 4 The Measurement report results for evaluation criteria of projects

Criteria	Observed Average	Fair Average	Model		Infit		Outfit	
			Measure	Error	Square Average	Z	Square Average	Z
crt8	1,89	1,87	1,77	0,14	1,39	2,6	1,33	2,2
crt5	2,47	2,48	0,96	0,14	0,78	-1,7	0,78	-1,6
crt4	2,93	2,95	0,29	0,15	0,7	-2	0,7	-1,9
crt2	3	3,02	0,17	0,15	0,68	-2	0,68	-2
crt7	3,08	3,1	0,02	0,16	1,38	2	1,46	2,3
crt6	3,19	3,21	-0,19	0,17	0,6	-2,6	0,61	-2,5
crt1	3,24	3,25	-0,27	0,17	1,29	1,5	1,33	1,7
crt10	3,25	3,27	-0,3	0,17	0,96	-0,1	1,03	0,2
crt9	3,56	3,57	-1,06	0,21	1,1	0,5	1,09	0,5
crt3	3,65	3,66	-1,39	0,23	1,2	0,9	1,11	0,5
Average	3,03	3,04	0	0,17	1,01	-0,1	1,01	-0,1
Standard Deviation	0,49	0,5	0,86	0,03	0,29	1,8	0,29	1,8

Model, RMSE .17 Adj (True) S.D. .89 Separation 5.23 Strata 7.31 Reliability .96
Model, Fixed (all same) chi-square: 260.0 d.f.: 9 significance (probability): .00
Model, Random (normal) chi-square: 8.7 d.f.: 8 significance (probability): .37

According to Table 4, it is seen that the criteria which are the weakest points for students while preparing their projects were number 8 (defining the common effect expected from the project), number 5 (revealing the authentic value of the project) and number 4 (forming the theoretical frame of the project). The students were most comfortable with criteria 3 (determining key words found in the project abstract), 9 (creating the general budget of the project) and 10 (justifying the project budget).

From Table 4, it can be said that the findings are in compliance with the chi-square results testing whether "there is a meaningful difference between the difficulty of criteria used in the project evaluation" hypothesis about the constant effect and the separation index of 5.23 and the reliability index of 0.96 ($\chi 2 = 260.0$, $p < .05$). Thus, the absence hypothesis was rejected, and it can be concluded that there is a statistically meaningful difference in terms of the difficulty/easiness of the criteria used in evaluation of the students' projects.

Rater-Project Bias Analysis

The rater-project bias analysis results are given in Table 5 to examine the presence of bias in the project evaluation performed by raters. According to this, the fact that the t values are in the range of -2 and +2 indicates there is an interaction bias (Linacre, 2014). Hence, it can be said that the t values range is from -3.30 to 2.07 and that some of the judges were biased. According to the findings, judge 2 was generous with project 2 and gave 36 points instead of 31 points, yet the same judge gave 36 points instead of 32 points for project 5. Similarly, judge 12 was severe with project 3 and gave 26 points instead of 31 points, but judge 9 was severe and gave 24 points instead of 32 points for project 2.

Table 5 Judge-project bias analysis

Observed Score	Expected Score	Number of Observation	Observed-Expected Average	Bias Model		Infit Squares Average	Outfit Squares Average	Judge	Project
				Size	Standard Error				
24	31,53	10	-0,75	-1,3	0,39	0,4	0,4	judge9	p2
26	30,71	10	-0,47	-0,82	0,4	0,7	0,7	judge1	p3
36	32,26	10	0,37	1	0,59	1	1,4	judge2	p5
36	31,21	10	0,48	1,22	0,59	0,7	1,1	judge2	p2
30,3	30,25	10	0	0,03	0,45	1	1	Average	
3,0	2,17	0	0,21	0,43	0,05	0,5	0,5	Standard Deviation	

Fixed (all = 0) chi-square: 61.2 d.f.: 72 significance (probability): .81

It is necessary to examine in detail the reasons why the judges were severe or generous for some projects and not for others.

CONCLUSION AND DISCUSSION

In this study, scientific research projects prepared by pre-service preschool teachers were examined in terms of several criteria according to MFRM (Many-Facet Rasch Model). In this analysis, the projects of the students, severity/leniency of judges' scoring, and consistency of the criteria defined were examined using MFRM. According to the findings, out of P6 projects, project P5 was the most successful project. However, project P1 was considered as the least successful project. The total score obtained for project P5 was 381 while the total score for project P1 was 349. It was revealed that among the judges, the most generous scoring was given by judge J11, and the most severe scoring was that of judge J4.

Another result obtained from the research was whether the criteria were easy or not for the students. When the findings obtained according to MFRM are taken into consideration, it was concluded that the easiest was criteria 3, “Writing Key Words Found in the Project Abstract”. Key words define the words used in the main title of the research project. In addition, key words are also defined as important words related to the topic selected for the project conducted by the researcher. Additionally, these key words also enable other researchers to access similar research projects (Bell, 2010; Lester and Lester, 2015). In our study, it can be said that in comparison with other criteria, the students did not have any difficulties in finding key words suitable for their projects during the preparation of their project proposal. Nevertheless, it was revealed that criteria 8, “Revealing the Common Effect Expected from the Project”, was the most difficult criteria for students during preparation of the project proposal. It can be said that students had difficulties in expressing the common effect of the project proposal at the writing stage and that they could not convey the common effect successfully. The common effect of the project is an important component of the project proposal. The researcher should clearly express the proposed research project’s originality, productivity, and contribution to society under the common effect heading (TÜBİTAK, 2014). In other words, the common effect is the researcher's consideration of the project's contribution to scientific knowledge, the economy and public welfare upon conducting the project (TÜBİTAK, 2018). Another important component in the common effect is to make a contribution to the project shareholders and projects to be realized in the future (TÜA, 2016). It is seen that students cannot enunciate the common effect of the research project at the targeted level. This may result from failings of the academic member teaching the course. On the other hand, this can also be due to the fact that students do not know what the common effect constitutes as a concept, and they did not conduct enough research to learn its meaning.

In studies conducted on the MFRM (Baştürk, 2009; Baştürk and Işıkoğlu, 2007; Batdı, 2017; Batdı and Elaldı, 2016; Köse, Usta and Yandı, 2016), rater bias emerged as a crucial factor and it is of great importance in terms of the reliability and validity of the results. According to the rater bias results obtained in this study, it was seen that the judge J2 gave 36 points instead of 31,21 and gave a generous scoring for project P2. On the other hand, judge J9 gave 24 points instead of 31,53 and severe scored project P2. When the reliability coefficients of the study were examined they were calculated as between 0.70 and -.0.96. It is possible to state that there was no problem in terms of the reliability of the analysis conducted and that it was reliable at a good level (Şencan, 2005).

RECOMMENDATIONS

Proposals based on the research results may be summed up as follows:

1. It was found that students have difficulties with some of the project preparation processes (e.g. expressing the common effect). Accordingly, every project preparation stage should be given more attention by the students in the project preparation process and the project proposal should be prepared by spending more time on its application.

2. According to the study's results, it was found that some judges were biased against some projects during the evaluation stage. Therefore, a short training course should be provided for judges to explain how the scoring should be conducted.

3. The study results were based on only quantitative data. The subjects which the students struggled with most, and the students' views, could be usefully examined after obtaining the quantitative results.

Proposals for future studies:

1. Only the facets regarding the criteria, project and judges were taken into consideration in the scope of the current study. Students could be added to the study as the fourth facet. Thus, the project preparation process of the students could also be investigated individually.

2. The prepared project proposals of the students could also be analyzed with different samples. Consequently, it would be possible to recognize whether there is any bias by the judges, using the same criteria with different samples.

ACKNOWLEDGEMENTS OR NOTES

This study is supported by Canakkale Onsekiz Mart Univesity Scientific Research Projects Unit (Project Number: SBA-2019-2920). We would like to thank Canakkale Onsekiz Mart Univesity Scientific Research Projects Unit for this support

REFERENCES

Akar, Ü. (2007). Ögretmen adaylarının bilimsel süreç becerileri ve elestirel düsünme beceri düzeyleri arasındaki ilişki, Yayımlanmamış Yüksek Lisans Tezi, Afyonkocatepe Üniversitesi Sosyal Bilimler Enstitüsü Sınıf Ögretmenliği Bilim Dalı, Afyon.

Akgün, S. (2000). Ögretmen ve ögretmen adaylarına fen bilgisi ögretimi. 6. Baskı. Pegem A Yayıncılık.

Akkanat, Ç., Abu, N.K., & Çakır, R. (2017). Öğretmen adaylarının bilimsel araştırma yöntemleri dersindeki motivasyonel inançları ve öğrenme stratejilerinin çeşitli değişkenler bakımından incelenmesi, Hasan Ali Yücel Eğitim Fakültesi Dergisi, 14 (27), 223-244.

Aksu, C. (2018). Müzik öğretmeni adaylarının 'bilimsel araştırma yöntemleri dersi'ne dönük görüşleri, Kastamonu Eğitim Fakültesi Dergisi, 26 (6), 2127-2138.

Ary, D., Jacobs, L.C., & Razavieh, A. (2006). Introduction to research in education (5th ed.). Fort Worth, TX: Harcourt Brace.

Aydoğdu, B. (2009). Fen ve teknoloji dersinde kullanılan farklı deney tekniklerinin öğrencilerin bilimsel süreç becerilerine, bilimin doğasına yönelik görüşlerine, laboratuvara yönelik tutumlarına öğrenme yaklaşımlarına etkileri, Yayımlanmamış Doktora Tezi, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, İlköğretim Anabilim Dalı Fen Bilgisi Öğretmenliği Programı, İzmir.

Bağcı-Kılıç, G., Haymana, F. & Bozyılmaz, B. (2008). İlköğretim fen ve teknoloji dersi öğretim programi'nin bilim okuryazarliği ve bilimsel süreç becerileri açisindan analizi. Eğitim ve Bilim, 33 (150), 52-63.

Baker, F.B. (2001). The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Bastürk, R. & Işıkoglu, N. (2008). Analyzing process quality of early childhood education with many facet rash measurement model. Educational Sciences: Theory and Practice, 8(1), 25–32.

Baştürk, R. (2009). Applying the many–facet Rasch model to evaluate powerpoint presentation performance in higher education. Assessment and Evaluation In Higher Education, 33(4), 431–444.

Batdı, V. (2017). Comparing the high school english curriculum in turkey through multi-analysis. Educational Sciences: Theory and Practice 17(4), 1255-1290.

Batdı, V., & Elaldı, Ş. (2016). Analysis of high school english curriculum materials through rasch measurement model and maxqda. Educational Sciences: Theory & Practice, 16, 1325–1347.

Best, J.W. & Kahn, J.V. (2017) Eğitimde araştırma yöntemleri, nitel araştırma bölümü (Çeviren: Mustafa Durmuşçelebi), (Editör: Onur Köksal). Konya: Dizgi Ofset.

Bins, I.C. (2009). Representatıon of scientific methodology in secondary science textbooks, Doctoral dissertation, University of Virginia, USA.

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2008). Scientific research method. Ankara: Pegem A Press.

Çelik, P. (2013). Probleme dayalı öğrenmenin öğretmen adaylarının fizik dersi başarısı, öğrenme yaklaşımları ve bilimsel süreç becerileri üzerindeki etkisi, Yayımlanmamış Doktora Tezi, T.C. Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü Ortaöğretim Fen ve Matematik Alanlar Eğitimi Anabilim Dalı Fizik Öğretmenliği Programı, İzmir.

Cengiz, C. & Karataş, F.Ö. (2014). Alan eğitiminde araştırma projesi dersi: biyoloji öğretmen adaylarının yansıtmaları [Introduction to Education Research Course: Reflections from Pre-service Biology Teachers]. Eğitim ve Öğretim Araştırmaları Dergisi, 3(4), 51-58.

Chapman, J.E., McCart, M.R., Letourneau, E.J., & Sheidow, A.J. (2013). Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. Journal of Consulting and Clinical Psychology, 81(4), 674.

Creswell, J.W. (2014). Nitel, nicel ve karma yöntem yaklaşımları araştırma deseni. (Çev. Ed. Demir, S. B.). Ankara: Eğiten Kitap.

De Ayala, R.J. (2009). The Theory and practise of item response theory. Series Editor Note By David. A. Kenny. The Guilford Press: New York, London.

Ekiz, D. (2009). Bilimsel araştırma yöntemleri (Genişletilmiş 2.Baskı). Ankara: Anı Yayıncılık.

Ersoy & İncebacak (2016). Okulöncesi öğretmen adaylarının bilimsel araştırmaya bakışları, Qualitative Studies (NWSAQS),11(4), 67-77.

Eti, İ. & Gündoğdu, S. (2016). Okul öncesi öğretmenliği dördüncü sınıf öğrencilerinin araştırma projesi dersine ilişkin görüş ve önerileri. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 16 (1), ss:121-139.

Gall, M.D., Borg, W.R., & Gall, J.P. (2002). Educational research: an introduction (7th ed.). Old Tappan, NJ: Pearson Allyn & Bacon.

Garza, A.L. (2015). Revisioning scientific methodologies and epistemologies with feminism and indigenous understandings, Master Thesis, Department of Women's Studies, College of Arts And Sciences, Denton, Texas, USA.

Güler, N. (2008). Klasik test kuramı genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma. Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Haiyang, S. (2010). An application of classical test theory and many facet Rasch measurement in analyzing the reliability of an English test for non-English major graduates. Chinese Journal of Applied Linguistics, 33(2), 87-102.

Hambleton, R.K. (1995). Meeting the measurement challenges of the 1990s and beyond: New assessment models and methods. In T. Oakland & R.K. Hambleton (Eds.), International perspectives on academic assessment (pp. 83-104). Boston, MA: Kluwer Academic Publishers.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). Fundamentals of item response theory. Newbury Park, CA: SAGE Publications, Inc.

Herman, B. C., Clough, M. P., & Olson, J. K. (2013). Teachers' nature of science implementation practices 2-5 years after having completed an intensive science education program. Science Education, 97(2), 271-309.

Hypolite, K.L. (2012). Understanding the views of the nature of science of undergraduate science, technology, engineering, and mathematics students, Doctoral dissertation, The faculty of the Graduate School Southern University and A & M College, United States.

İlhan, A, Çelik, H.C., & Aslan, A. (2016). Üniversite öğrencilerinin bilimsel araştırmaya yönelik tutumlarının incelenmesi. İnönü Üniversitesi Eğitim Fakültesi Dergisi, 17(2), 141-156. doi: 10.17679/iuefd.17218132.

İlhan, M. (2015). Standart ve solo taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeyli rasch modeli ile incelenmesi. Doktora Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

İlhan, M. (2016). Açık uçlu sorularla yapılan ölçmelerde klasik test kuramı ve çok yüzeyli Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılması. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 31(2), 346-368. doi:10.16986/HUJE.2016015182.

Johnson, B., & Chrıstensen, L. (2014). Eğitim araştırmaları: nicel, nitel ve karma yaklaşımlar. (Çev. edt: Demir, S.B.) Ankara: Eğiten Kitap.

Jöreskog, K.G. & Sörbom, D. (1993). Lisrel: User's reference Guide, Chicago: Scientific International Software.

Kaya, G., & Yılmaz, S. (2016). Açık sorgulamaya dayalı öğrenmenin öğrencilerin başarısına ve bilimsel süreç becerilerinin gelişimine etkisi. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 31 (2), 300-318.

Kefi, S., Çeliköz, N. & Erişen, Y. (2013). Ökulöncesi eğitim öğretmenlerinin temel bilimsel süreç becerilerini kullanım düzeyleri. Eğitim ve Öğretim Araştırmaları Dergisi, 2(2), 300-319.

Kim, Y., Park, I., & Kang, M. (2012). Examining rater effects of the TGMD-2 on child. Adapted Physical Activity Quarterly, 29, 346-365.

Köse, İ. A., Usta, H.G. & Yandı A. (2016). Sunum yapma becerilerinin çok yüzeyli rasch analizi ile değerlendirilmesi. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 16(4), 1853-1864.

Lawshe, C.H. (1975). A quantitative approach to content validity. Personnel Psychology, 28, 563–575.

Linacre, J.M. (2003). Size vs. significance: Standardized chi-square fit statistic. Rasch Measurement Transactions, 17(1), 918.

Linacre, J.M. (1989). Many-facet Rasch measurement. Chicago: MESA Press.

Linacre, J.M. (2014). A user's guide to facets rasch-model computer programs. 30 Aralık 2014 tarihinde http://www.winsteps.com/a/facets-manual.pdf adresinden alınmıştır

Mahoney, S. (2013). educational research in the united states: a survey of pre-k-12 teachers' perceptions regarding the purpose, conceptions, use, impact, and dissemination. Doctoral dissertation, Arizona State University, Phoenix, Arizona, USA.

MEB (2017). Öğretmenlik mesleği genel yeterlilikleri. Ankara: Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü.

MEB. (2015). Fen Bilimleri dersi öğretim programı (ilkokul ve ortaokul 3, 4, 5, 6, 7 ve 8. sınıflar). Ankara: Milli Eğitim Bakanlığı Yayınları.

Mulqueen C., Baker D., & Dismukes, R.K. (2000, April) Using multifacet Rasch analysis to examine the effectiveness of rater training. Presented at the 15th Annual Conference for the Society for Industrial and Organizational Psychology. New Orleans. 15 Mart 2017 tarihinde http://www.air.org/files/multifacet_Rasch.pdf adresinden alınmıştır.

Murray, K. (2017). The priorıty of democracy to education research, Doctoral dissertation, University of Colorado Boulder, USA.

O'Neill, T.R., & Lunz, M.E. (1996, April). Examining the invariance of rater and project calibrations using a multi-facet Rasch model. Paper Presented at the Annual Meeting of the American Educational Research Association, New York.

Orçan, B. (2013). Lisansüstü eğitim programlarında öğrenim gören öğrencilerin bilimsel araştırma sürecindeki kavram yanılgılarının belirlenmesi, Yayımlanmamış Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı, Ankara.

Patton, Q.M. (2014). Nitel araştırma ve değerlendirme yöntemleri. Mesut Butun & Selcuk Besir Demir (Trans.Eds.). Ankara: PegemA Publishing.

Plano Clark, V. L., & Creswell, J. W. (2010). Understanding research: A consumer's guide. Upper Saddle River, NJ: Pearson Education.

Robson, C. (2015). Bilimsel araştırma yöntemleri. (Ed. Ş. Çınkır & N. Demirkasımoğlu). Gerçek Dünya Araştırması (1. Basım). 97-134. Ankara: Anı Yayıncılık.

Schwartz, R.S., Westerlund, J., Garcia, D., & Taylor, T. (2010). The impact of full immersion scientific research experiences on teachers' views of the nature of science. Electronic Journal of Science Education, 14(1), 1–40.

Şencan, H. (2005). Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik. Ankara: Seçkin Yayıncılık.

Spang, E. J. (2008). How does teacher education make a difference in our schools? beginning science teachers' and their students' and use of scientific inquiry, Doctoral dissertation, The School of Education of Stanford Unıversıty, USA.

Sudweeks, R.R., Reeve, S., & Bradshaw, W.S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. Assessing Writing, 9(3), 239-261.

TUA (2016). Türkiye Cumhuriyeti Avrupa Birliği Bakanlığı eğitim ve gençlik programları merkezi bakanlığı Türkiye ulusal ajansı, Adım Adım Proje Geliştirme Fikri El Kitabı, Ankara: Eurodesk Türkiye Eğitim Kitleri Seti.

TÜBİTAK (2014). Türkiye bilimsel ve teknolojik araştırma kurumu, 45. ortaöğretim öğrencileri araştırma projeleri yarışması, Proje Rehberi, Ankara: TÜBİTAK yayınları.

TÜBİTAK (2018). Türkiye bilimsel ve teknolojik araştırma kurumu 1001 – bilimsel ve teknolojik araştırma projelerini destekleme programı proje başvuru formu [http://tubitak.gov.tr/tr/destekler/akademik/ulusal-destek-programlari/1001/icerik-basvuru-formlari] web adresinden 5.3.2019 tarihinde indirildi.

Wright, B.D., & Linacre, J.M. (1994). Reasonable mean-square fit values. Rasch Measurement: Transactions of the Rasch Measurement SIG, 8(3), 370.

Yaşar, M. (2014). Bilimsel araştırma yöntemleri dersine yönelik tutum ölçeği geliştirme çalışması: geçerlik ve güvenirlik. Journal of Educational Sciences Research, 4(2), 109-129. http://dx.doi.org/10.12973/jesr.2014.42.7

YÖK (2007). Öğretmen yetiştirme ve eğitim fakülteleri, Ankara: T.C. Yüksek Öğretim Kurulu.

YÖK (2018). Yeni öğretmen yetiştirme lisans programları. [https://www.yok.gov.tr/kurumsal/idari-birimler/egitim-ogretim-dairesi/yeni-ogretmen-yetistirme-lisans-programlari] web adresinden 04.03. 2018 tarihinde indirildi.

Yue, X. (2011). Detecting rater centrality effect using simulation methods and Rasch measurement analysis. Unpublished Doctoral Thesis, Virginia State University, Petersburg.

Yurdugül, H. (2005, Eylül). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. XIV. Ulusal Eğitim Bilimleri Kongresi'nde sunulmuş bildiri, Pamukkale Üniversitesi, Denizli.

Zientek, L.R., Capraro, M. & Capraro, R. (2008). Reporting practices in quantitative teacher education research. Educational Researcher, 37(4), 208-216