Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 345-352, Kasım 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 345-352, November 2020

Research Article

Rasch-based Assessment for Innovative Education Systems

Elif Dülger

Sakarya University, Engineering Faculty, Industrial Engineering Department, Sakarya, Turkey (ORCID: 0000-0002-8162-8726)

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2020 – 22-24 October 2020)

(**DOI:** 10.31590/ejosat.824163)

ATIF/REFERENCE: Dülger, E. (2020). Rasch-based Assessment for Innovative Education Systems. *European Journal of Science and Technology*, (Special Issue), 345-352.

Abstract

Generating high-quality questions is one of the most significant things to improve learning process quality. One of the contemporary methods of developing high-quality questions is the Rasch Method because of superiority for scale development. The most important advantage of the Rasch method is that it allows the measurement of item calibration which is independent of a person's ability. At the same time, it allows the measurement of a person's ability which is independent of item difficulty. In this way, instructors can criticize the quality of their questions, and educational institutions can update their own assessment systems for higher quality and innovative education. This study aims to be a guide that will meet this need of instructors based on the Rasch method. This approach makes it possible for well-qualified items can be used throughout the learning process to improve the student's learning experience. Thus it is possible to realize the most appropriate content and test design which is necessary for personalized education. On the other hand, it can analyze items psychometric properties and can improve the quality of evaluation systems for the continuous improvement system of future educational institutions. This study evaluated the properties of mathematics items within the test which is applied during term using the Rasch Model. The test was developed for use within a math course which used a modular approach based on learning objects. Results have been evaluated with a view to making recommendations for the improvement of the test psychometric properties. This approach not only enables course developers to evaluate the quality of their test in term but also provides useful information about their own process performance during the e-learning process and their own ability. Due to enable analyzing the student's ability while the e-learning process ongoing, it is possible to correct one's deficiencies. In this study, answers given to 20 items, provided by 62 students of the Internet and Network Technologies distance education program at the level of the associate degree in the fall semester of the 2019-2020 academic year, were examined. Measurement reports, fit statistics, reliabilities were calculated by the Winstep package program.

Keywords: Rasch Model, Assessment Quality, Learning Quality, Distance Education Quality, Education 4.0

Yenilikçi Eğitim Kurumları İçin Rasch Metodu Temelli Değerlendirme

Öz

Yüksek kaliteli sorular üretmek, öğrenme süreci kalitesini artırmak için en önemli şeylerden biridir. Yüksek kaliteli sorular geliştirmenin çağdaş yöntemlerinden biri, ölçek geliştirmedeki üstünlüğü nedeniyle Rasch Yöntemidir. Rasch yönteminin en önemli avantajı, kişinin yeteneğinden bağımsız olarak madde kalibrasyonunun ölçülmesine izin vermesidir. Aynı zamanda bir kişinin yeteneğinin madde zorluğundan bağımsız olarak ölçülmesini sağlar. Bu şekilde, eğitmenler sorularının kalitesini eleştirebilir ve eğitim kurumları daha kaliteli ve yenilikçi eğitim için kendi değerlendirme sistemlerini güncelleyebilir. Bu çalışma, öğretim elemanlarının bu ihtiyacını Rasch yöntemine göre karşılayacak bir rehber olmayı amaçlamaktadır. Bu yaklaşım, öğrencinin öğrenme deneyimini iyileştirmek için iyi nitelikli öğelerin öğrenme süreci boyunca kullanılabilmesini mümkün kılar. Böylelikle kişiselleştirilmiş eğitim için gerekli olan en uygun içerik ve test tasarımını gerçekleştirmek mümkündür. Öte yandan, öğelerin psikometrik özelliklerini analiz edebilir ve gelecekteki eğitim kurumlarının sürekli iyileştirme sistemi için değerlendirme sistemlerinin kalitesini iyileştirebilir. Bu çalışmada, Rasch Modeli kullanılarak dönem boyunca uygulanan matematik testi maddelerinin özellikleri değerlendirilmiştir. Test, öğrenme nesnelerine dayalı modüler bir yaklaşım kullanan bir matematik dersinde kullanılmak üzere geliştirilmiştir. Sonuçlar, testin psikometrik özelliklerinin iyileştirilmesine yönelik önerilerde bulunmak amacıyla değerlendirilmiştir. Bu yaklaşım, kurs geliştiricilerin testlerinin kalitesini dönem içinde değerlendirmelerini sağlamakla kalmaz, aynı zamanda e-öğrenme sürecinde kendi süreç performansları ve kendi yetenekleri hakkında faydalı bilgiler sağlar. E-öğrenme süreci devam ederken öğrencinin becerisinin analiz edilebilmesi sayesinde eksikliklerin giderilmesi mümkündür. Bu araştırmada 2019-2020 eğitim-öğretim yılı güz yarıyılı içerisinde, önlisans düzeyinde 62 İnternet ve Ağ Teknolojileri uzaktan eğitim programı öğrencisinin verdiği 20 maddeye verilen cevaplar incelenmiştir. Ölçüm raporları, uyum istatistikleri, güvenilirlikler Winstep paket programı ile hesaplanmıştır.

Anahtar Kelimeler: Rasch Model, Değerlendirme Kalitesi, Öğrenme Kalitesi, Uzaktan Eğitim kalitesi, Eğitim 4.0.

1. Introduction

The contemporary approach to education, Education 4.0, has an important position in today's digitalizing education system with the aim of maximizing the use of information, technology and internet. Countries have seen the strategic importance of contemporary e-learning systems during the Covid-19 pandemic. Only countries with this technological infrastructure could continue to implement their education curricula and thereby could improving their society, human capital. We can summarize main transformations in education area as follows (Dulger and Sertvuran, 2019): There have been four major transformation in education up to the present day as a result of the revolutions (Puncreobutr, 2016; Harkins, 2008). Education 1.0 is aimed at meeting the needs of agricultural society, Education 2.0 is for industrial society, Education 3.0 is for technology society and Education 4.0 is for meeting the needs of innovation society. Oztemel (2018) summarized the main points of change and transformation that will occur within the scope of Education 4.0 in his study evaluating new trends in education as follows: Independence from time and space, personalized education, flexible learning, blended learning, project-based learning, BYOD, time management, field information and experience, big data analysis, situation assessment instead of exams, collaboration with students in content development, virtual mentors, web interfaces, access systems and other Industry 4.0 technologies such as cloud computing, advanced robotics, bitcoin and blockchain (intelligent contracting, information security), autonomous devices, 3D printing (Oztemel, 2018).

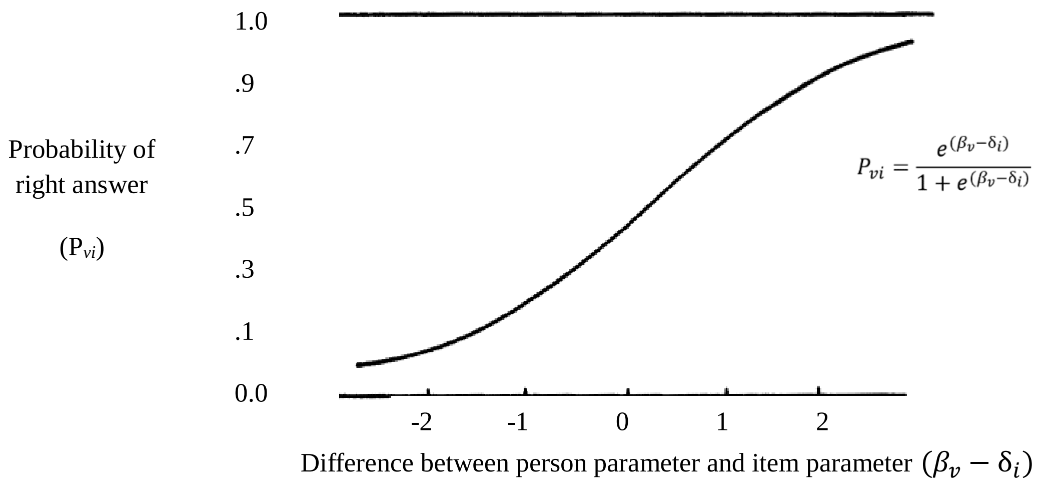
The indicators of the quality of e-learning systems is still a controversial issue. Even though educational technologies are well used, learning is related to the cognitive skills of the individual, and educators should be a facilitator for this process rather than transferring information in today's education systems. Managing this process requires monitoring the process and this can only be objectively interpreted by following the digital traces of the e-learning students and based on their performance from the learning test results. As one of the prominent names in measurement and evaluation in education systems, Dylan Wiliam says high-quality questions may be the most significant thing we can do to improve the quality of student learning (Wiliam, 2011). Increasing the quality of learning with quality questions means obtaining better learning outcomes, in other words, increasing the gain of learning outcomes. One of the views confirming this is expressed in the OECD report as follows: learning outcomes which is targeted by quality assurance are indeed the key to meaningful education (OECD, 2012). At that point, identifying and overcoming the learning problems on-time, are important to complete the learning process successfully and acquisition of learning outcomes. If studies on the assessment of learning are examined, it is seen that Assessment For Learning (AFL) has an important potential to improve learning. Because assessment has a very strong theoretical and experimental infrastructure not only as an evaluation of success but also in terms of including it in teaching process and increasing student participation (Wiliam, 2011). The study firstly aims to improve item quality for a better assessment system, then improve the e-learning quality through monitoring the e-learning process with quality items, thereby the institution quality.

2. Material and Method

2.1. Item Response Theory and The Rasch Method

Item Responce Theory (IRT) is a theory based on mathematical model that reveal the relationship between individuals' abilities that cannot be directly observed and their responses to test items prepared to measure their abilities (İlhan, 2016). This mathematical model can define the examinee's ability independently from the test. The focal point is not results that reached at end of the test, like Classical Test Theory, the focal point is the ability level estimation of the examinee. According to this theory, the estimated ability level is the same even if two different tests including different questions are applied (Ozyurt et al., 2012). IRT consists of three different models. Single parameter logistic Model (1PL Model) known as the Rasch Model and consisting of item difficulty parameter, Two Parameter Logistic Model (2PL Model) consisting of item difficulty and item discrimination parameters and Three parameter Logistic Model (3PL model) consisting of item difficulty, item discrimination and pseudo-chance parameters (Finch 2008; Battisti et al, 2005).

The Rasch Model assumes that the probability of any person to choose a category in any item is equal to the logarithm of the difference between the person's ability level and the item difficulty, and this model is a mathematical relationship that takes advantages of interaction between people and items. One of the most important advantages of this model is that it allows the measurement of item calibration which is independent of a person's ability. At the same time it allows the measurement of a person's ability which is independent of item difficulty. This advantage enables a generalization beyond the sample and the Rasch measurement model has been successfully applied to testing since 1965 with large scale implementations. (Linacre 1989; Koparan and Guven 2014). The model expresses the probability of the right response by the following relation (Wright, 1977) in Figure 1.



δ_i ; the difficulty of item$_i$

βv ; the ability of the person$_v$

Figure 1. The Rasch Model Characteristic Curve

The Rasch model makes possible easier analyze with particular unit measure called logit (Wright and Stone, 1979) and through the instrument of logitmap, it can be interpreted item difficulty and person ability on the same map. Rasch analysis indicate the degree to which the analysis reliability distinguishes the level of quality among the items using the reliability statistics. The analysis also helps to identify, through fit statistics, any item that do not fit the model and any candidate whose scores do not appear to be consistent with the model (Basturk, 2008).

One of the prominent studies related to Rasch Method in Turkish Higher Education System was carried out by R. Basturk (Basturk 2008). The study investigated the usefulness of the Many-Facet Rasch Model (MFRM) in evaluating the quality of performance related to PowerPoint presentations on science education students. The paper specifically investigated presentation ability in terms of item/task difficulty and rater severity/leniency. As a result, the study revealed that analysing the outcomes of peer assessment through MFRM techniques provided rich approaches suitable for assessing student performance in higher education. When we look at other prominent studies using Rasch Method in higher education, it can be said that general common point is its powerfulness to evaluate and improve education. The Rasch Method could provide valuable information on test reliability, item difficulty and examinee ability to evaluate higher education students' performance and items' quality. These studies have revealed the usefulness of The Rasch Method with its applications in their studies in different areas such as nursing, engineering, and medical science. Some of these studies used Extended Logistic Model of Rasch, some of them were used for the improvement of teaching methods.

Waugh (1999) aimed to create a new range level scale, in its "Approaches to studying for students in higher education: A Rasch measurement model analysis" work. He analysed it's psychometric properties with the Extended Logistic Model of Rasch using the sample consisted of 369 students from an Australian university. The data were analysed initially with the whole sample for the 40 attitude items and for the 40 behavioural items separately. Items not fitting the model were then discarded (12 attitude items and 12 behavioural items). The analysis was repeated using the 56 valid items together and the items of each of the five subscales separately. This subscales were Deep Approach, Surface Approach, Strategic Approach, Lack of Direction in Studying and Academic Self-Confidence towards Studying. Ultimately, it has been fifty-six of the 80 items formed a good scaled with satisfactory psychometric properties. This paper revealed that The Extended Logistic Model of Rasch was useful in creating an interval level scale of student attitudes and behaviours towards studying, and for analysing its psychometric properties and conceptual design. And Rasch model helps to explain how behaviours are influenced by attitudes (Waugh 1999).

Hagquist et al. (2009) used Rasch model to examine the psychometric properties of the nursing self-efficacy (NSE) scale. Data were collected among nursing students in Sweden. Two sets of items were analysed more thoroughly: an original set of nine items with eleven response categories and a revised set of seven items with seven response categories. Invariance of the item functioning and the categorisation of the items were analysed. Targeting was examined by comparisons of the items and persons locations. Differential Item Functioning across sample groups such as gender was examined using analysis of variance. The final set of seven items was also analysed more closely with respect to possible multidimensionality and response dependence. Consequently The Rasch analysis of the original set of nine items showed high reliability measured by a person separation index, but it also indicated severe problems with the targeting, the categorisation of the items as well as lack of invariance. Although the revised set comprising seven items with seven categories performed better than the original item set some items showed misfit according to formal test statistics. Graphical examination showed, however, that the items operated in the right direction. The formal test of local independence of the items indicated minor signs of multidimensionality, alternatively response dependence. This study revealed that The Rasch model is useful for rigorous examination and development of measurement instruments in nursing research and facilitates disclosure of lack of invariance and other measurement problems that may not be easily detected by traditional analyses. (Hagquist et al., 2009).

Ghulman and Mas'odi (2009) developed assessment model based on Rasch Measurement Model which can be used to improve the students' assessment method and validate the CLO (Course Learning Outcomes) of each course. At the same time this study provides an overview of an alternative "modern" measurement as practiced using item response theory with focus on Rasch measurement model. A case study in the College of Engineering, Umm al-Qura', Makkah on students (N=75) was conducted to measure their command of knowledge (Kn) and understanding (uN); as categorised according to Bloom's Taxonomy. The test for Manufacturing Process III course was administered on 3rd year students from the Department of Mechanical Engineering, Umm al_ Qura' University. The result from the test were tabulated and run in WinSteps, a Rasch Analysis software. Generally the students fared poorly below the expected performance achieving a poor mean; μperson mean of only 42.80% which is way below the 60% threshold limit Level of Learning Competence. This study indicates that the undergraduates encounter significant difficulties in grasping some fundamental engineering principles in extrusion. The study showed that SPELA as a model of measurement can provide better estimate of students' ability more accurately based on the CLO as compared to the traditional Cumulative Grade Point Average (CGPA) method of assessment using raw score (Ghulman and Mas'odi, 2009).

Yang et al. (2011) used the Rasch model to analyse an examination in anaesthesiology for medical students to identify examinee ability and appropriateness of the test content. Fifty items has been administered to 119 fifth- and sixth-year medical students in the exam. The Rasch model was used to perform item analysis of the examination. Misfit items or examinees were excluded first, then test reliability was assessed with reliability indices. After the exclusion of two misfit items and one misfit examinee, the estimated test reliability was only 0.63. The mean item difficulty was set at 0 by definition (SD = 2.02) and the mean examinee ability was 1.56 (SD = 0.71), which means that the examinees were able to correctly answer 83% of items on average. This study has been demonstrated that statistical item analysis with the Rasch model could provide valuable information related to test reliability, item difficulty and examinee ability, which could be applied to further item modification and future test development of clinical curriculums for medical students. In conclusion, we demonstrated that the Rasch analysis could be applied satisfactorily to item analysis of the examination in anaesthesiology for medical students. It has been also provided valuable information for further item modification and future test development in clinical curriculums (Yang et al. 2011).

Ibrahim et al. (2012) examined physics examination paper using the Rasch model. The sample consists of 119 male and female pre-university students from the University of Malaya, Kuala Lumpur, who sat for the Basic Physics 3 paper in the session 2010/2011. Out of 119 students, 54 of them are from Physical Science Programme, and the rest of them are from the Biological Science Programme. For this study, the questions are classified according to Bloom's Taxonomy and every part of each question is labelled as a different item. The study has been suggested that the item model gives a good fit but persons and items measurement gives a fair fit. The measurement has been also indicated that a number of students encounter significant difficulties in grasping some of the fundamental topics in physics. These findings has been provided valuable insights into the structure and functionality of the examination paper. The study has been showed that the instrument construct for the final examination Basic Physics 3 is reliable and of good quality. The measurement has been conducted also reveals the true degree of cognitive learning abilities of 119 pre-University or Matriculation students in University of Malaya who sat for the final examination paper Basic Physics 3 in 2010. It has been also gave information determining the quality of the examination paper set. The authors expressed that the information generated from this measurement can be used as a guidance to determine the appropriate improvement of the teaching method (Ibrahim et al., 2012).

Researches shows that personalized e-learning system based on Item Response Theory estimates the abilities of e-learning students and consequently can used for the appropriate content and test design which has a significant role for personalized education (Chen et al., 2005).

2.2. Methodology

The methodology of the study rest upon a case study analysis of higher education two years of distance education in Turkey. The effectiveness of the mathematics e-course was evaluated via performance measurement of the distance education students at Sakarya University, Turkey. Firstly, the mathematics course was re-structured according to a modular approach and Shareble Content Objects and its test items because of using SCORM supported LMS of institution. Then, the performance of distance education students was measured. Item analysis previously has been completed with test results of Computer Programming programme (Dulger, 2012).

In this study, test items have been analysed with Rasch model with the aim of improving the tests to measure the students' learning and providing useful information to the students themselves and to those running the courses. The data consists of responses from Internet and Network Technologies Programme distance education students in the fall semester of the 2019-2020 academic year, so the data belongs to the 2019 year.

3. Findings and Results

3.1. Results

The logit scale can be seen on the left side of the figure under the Measure column. The negative logits denote easier items and persons of lower ability than the positive logits. The distribution of persons who took the assessment are displayed on the left side of the dashed vertical line and the items themselves on the right side of the line. The letter M denotes the mean person ability and mean item difficulty. The letter S denotes 1 standard deviation above and below the mean, and the letter T denotes two standard deviations.

We can see that the item difficulties appear to be well matched to the abilities of the students taking the test (in the range of ±2 standard deviations), the means and standard deviations of both are close together. The logitmap for the scores given by 62 students for the 20 items is shown in the Figure 2 below.

```
INPUT: 62 PERSON  20 ITEM  REPORTED: 62 PERSON  20 ITEM  2 CATS MINISTEP 4.7.0.0
--------------------------------------------------------------------------------

MEASURE PERSON - MAP - ITEM
         <more>|<rare>
    4  XXXXXX  +
               |
               |
               |
               |
              T|
               |
       XXXXXX  |
               |
    3          +
               |
               |
               |
               |
          XXX  |
               |
               |
               |
    2      XX  +
            X S|
               |T
               |  I0012
               |  I0013
               |
           XX  |
               |  I0011
               |
    1      XX  +
            X  |S
      XXXXXXX  |  I0010
            X  |
           XX  |
           XX M|  I0019
            X  |  I0020
               |  I0004  I0014  I0016  I0017
    0    XXXX  +M
            X  |  I0007  I0018
           XX  |  I0002
               |  I0001
           XX  |
               |  I0003
            X  |  I0005  I0009  I0015
           XX  |
           XX  |S
   -1          +  I0006
               |
           XX S|
            X  |
               |
         XXXX  |
               |
            X  |T
            X  |
   -2      XX  +
               |
               |  I0008
               |
            X  |
               |
               |
              T|
               |
   -3          +
         <less>|<freq>
```

Figure 2. Logitmap for mathematic test

The most difficult item was I12 and the easiest item was I8. But many of the questions are too easy for some of the students. For those students at the upper end of the ability range, there are too few questions to be able to reliably discriminate between their performances. Alternatively, it could be highlighting an issue that the teaching was not aligned with the needs of the students. The provision of the information from the Rasch analysis can be useful for course leaders and teachers who wish to examine the performance of their students, as well as the test developers. From a test development perspective, the test needs to contain several more difficult items if students' ability is to be measured more reliably. Table 1 shows the item analysis report.

Table 1. Item statistics

INPUT: 62 PERSON 20 ITEM REPORTED: 62 PERSON 20 ITEM 2 CATS MINISTEP 4.7.0.0

PERSON: REAL SEP.: 2.02 REL.: .80 ... ITEM: REAL SEP.: 2.21 REL.: .83

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
11	26	60	1.14	.34	1.28	1.62	1.97	2.59	A .50	.64	74.1	77.2	I0011
13	23	58	1.48	.36	1.23	1.22	1.83	2.01	B .55	.65	75.0	79.4	I0013
12	22	60	1.62	.36	.99	-.01	1.55	1.36	C .63	.66	85.2	80.9	I0012
4	35	62	.20	.32	1.10	.71	1.50	1.72	D .53	.59	69.6	74.0	I0004
14	35	60	.12	.33	1.25	1.65	1.39	1.34	E .48	.59	68.5	74.5	I0014
7	37	62	-.01	.32	1.13	.92	1.36	1.25	F .51	.57	69.6	74.1	I0007
10	30	62	.72	.33	1.24	1.56	1.31	1.17	G .53	.62	73.2	74.3	I0010
2	38	60	-.23	.33	.98	-.09	1.30	.97	H .56	.56	74.1	75.1	I0002
18	37	62	-.01	.32	.88	-.85	1.19	.73	I .60	.57	80.4	74.1	I0018
15	42	61	-.62	.34	1.14	.94	1.03	.22	J .48	.53	72.7	76.3	I0015
5	42	60	-.65	.34	1.04	.32	.85	-.28	j .51	.52	72.2	76.4	I0005
20	34	62	.30	.32	.99	-.05	.99	.03	i .60	.60	76.8	73.9	I0020
1	38	60	-.26	.33	.92	-.47	.74	-.80	h .60	.56	75.9	75.1	I0001
3	42	62	-.54	.33	.89	-.70	.75	-.67	g .58	.53	76.8	76.0	I0003
17	35	61	.11	.32	.86	-.99	.81	-.63	f .64	.58	81.8	73.8	I0017
6	45	61	-.97	.35	.82	-1.13	.67	-.69	e .56	.49	85.5	78.1	I0006
16	35	61	.19	.32	.82	-1.35	.67	-1.30	d .66	.58	80.0	73.7	I0016
9	43	61	-.68	.34	.77	-1.56	.62	-1.03	c .61	.51	87.3	76.6	I0009
8	54	61	-2.25	.44	.64	-1.49	.29	-.85	b .50	.35	89.1	87.5	I0008
19	33	61	.35	.33	.60	-3.18	.46	-2.47	a .76	.61	87.3	74.3	I0019
MEAN	36.3	60.9	.00	.34	.98	-.1	1.06	.2			77.8	76.3	
P.SD	7.4	1.0	.85	.03	.20	1.3	.45	1.3			6.3	3.2	

Rasch fit statistics means the fit of the items to the Rasch model. Lunz M., Wright B. and Linacre J. reported two kinds of fit statistics; infit and outfit. The infit statistic is the information on weighted mean-square residual difference between observed and expected, which focuses on the accumulation of central, inlying, deviations from expectation. The outfit statistic is the usual unweighted mean-square residual, which is particularly sensitive to outlying deviations from expectation. As a rule of thumb, Linacre suggested 0.50 as a lower-control limit and 1.50 as an upper-control limit for infit and outfit statistics. Also, some researchers suggested using a narrower range defined by a lower-control limit of 0.70 (or 0.75) and an upper-control limit of 1.30 (Eckes 2009).

Even though there is a minor difference between researchers, value limits mentioned in the literature for both the infit and the outfit statistics are rather similar (Guler, 2014). These characteristics are important for the test development process. Questions 4, 8, 11, 12, 13 and 19 display outfit statistics outside the acceptable values. The other outfit values and infit values are within acceptable limits. Hence, it can be said that there is no problem about items characteristics for test except this six questions.

For high test reliability, it must be person sample size with a large ability range and/or a test with many items. Since participation in the developed tests is not compulsory, person reliability range between .65 and .75 (low stakes tests) is adequate. It can be seen in Table 2 which shows that person reliability is .80 and item reliability is .83.

Table 2. Students' Measurement Report

Time for estimation: 0:0:0.161
Output to C:\Users\Sau\Desktop\ZOU016WS.TXT
MATH TEST

PERSON	62 INPUT		62 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	11.7	19.6	.76	.75	1.00	.0	1.06	.1
P.SD	5.6	.8	1.92	.40	.21	.9	.54	.9
REAL RMSE	.85	TRUE SD 1.72	SEPARATION 2.02		PERSON RELIABILITY .80			

ITEM	20 INPUT		20 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	36.3	60.9	.00	.35	.98	-.1	1.06	.2
P.SD	7.4	1.0	.85	.03	.20	1.3	.45	1.3
REAL RMSE	.35	TRUE SD .78	SEPARATION 2.21		ITEM RELIABILITY .83			

As the Rasch reliability coefficient approaches to +1.00, the reliability increases as it is in the traditional interpretation of reliability results. Thus, we can say that item and person reliability level is enough degree for the test.

4. Conclusions

The effects of the Industrial 4.0 revolution bring about the transformation of educational systems as in other fields. There is a need for technology-intensive restructuring of education systems in this period when everything is digitalized. It is an important requirement to develop an education management system that will focus on knowledge and skill-intensive competent individuals. On the other hand, education management systems should implement content design based on technological opportunities and blended with innovation culture at the same time. In this study, how to use these technologies in a beneficial way for the individual needs of the learner is discussed. Apart from these; for todays education systems, monitoring the e-learning process with quality items has a significant role for structuring the individual learning path according to the curriculum created individually, designing the course content in accordance with the learner specific needs and abilities, presenting the sufficient information students' needs when they need.

In this study, the mathematic performance of Sakarya University distance education students and the test items of the course were analyzed with the Rasch method. Rasch Method analyzes items psychometric properties and can use to improve the assessment quality which takes an important issue for the e-learning performance of distance education students. This study can be an example of the application of the Rasch model within the quality assurance system to enhance the quality of the e-learning process because of make possible an increase in the degree of learning outcomes acquisition. As a result of this study, the Rasch Method can be actively used in the evaluation of teaching and test design in higher education, and additionally, it can be used as an analysis tool in the development of the individual abilities of distance education students. This individual analysis has a significant factor for today's education systems due to the necessity of personalized education.

When we look at other prominent studies using Rasch Method in higher education, it can be said that general common point is its powerfulness to evaluate and improve education. The Rasch Method could provide valuable information on test reliability, item difficulty and examinee ability to evaluate higher education students' performance and items' quality. These studies have revealed the usefulness of The Rasch Method with its applications in their studies in different areas such as nursing, engineering, and medical science. Some of these studies used Extended Logistic Model of Rasch, some of them were used for the improvement of teaching methods (Waugh 1999; Hagquist, Bruce, and Gustavsson 2009; Ghulman and Mas'odi 2009; Yang et al. 2011; Ibrahim et al. 2012). In conclusion, the application of the Rasch model in the evaluation of the assessment used to measure students' performance offers an improvement to the quality assurance of the overall distance learning system. Thereby the aim was to improve the evaluation system quality of the institution mentioned above.

References

Baştürk R. (2008)." Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education", Assessment & Evaluation in Higher Education, 33:4, 431-444, DOI: 10.1080/02602930701562775

Battisti F.D., Nicolini G., Salini S. (2005), "The Rasch Model to measure service quality", Departmental Working Paper No. 27/2003 of Department of Economics, University of Milan, Italy. Retrieved from: http://wp.demm.unimi.it/files/wp/2003/DEMM-2003_027wp.pdf, Access date: 03.06.2020

Chen CM., Lee HM., Chen YH. (2005)."Personalized e-learning system using Item Response Theory", Computers & Education, Volume 44, Issue 3, April 2005, Pages 237-255, https://doi.org/10.1016/j.compedu.2004.01.006

Dülger E., (2012). "Yükseköğretimde uzaktan eğitim performansinin ölçme ve değerlendirilmesi için bir model önerisi", Doktora tezi, Sakarya University, Turkey

Dülger E., Sertvuran I. (2019). "Education 4.0: A system proposal based on lean learning", 10th International Symposium on Intelligent Manufacturing and Service Systems, Sakarya

Eckes, T. (2009) "Many-facet Rasch measurement." Section H in relating language examinations to the common european framework of reference for languages: Learning, Teaching, Assessment. Language Policy Division, Strasbourg

Finch H. (2008). "Estimation of item response theory parameters in the presence of missing data", Journal of Educational Measurement Fall 2008, Vol. 45, No. 3, pp. 225–245

Ghulman H. A., Mas'odi M. S. (2009) "Modern measurement paradigm in Engineering Education: Easier to read and better analysis using Rasch-based approach" International Conference on Engineering Education (ICEED 2009). Kuala Lumpur, Malaysia. December 7-8.

Güler, N. (2014) "Analysis of open-ended statistics questions with Many Facet Rasch Model." Eurasian Journal of Educational Research 55: 73-90

Hagquist, C.t, Bruce M., Gustavsson J.P. (2009) "Using the Rasch model in nursing research: An introduction and illustrative example." International Journal of Nursing Studies 46: 380-393. doi: 10.1016/j.ijnurstu.2008.10.007

Harkins, A.M. (2008). "Leapfrog Principles and Practices: Core Components of Education 3.0 and 4.0", Futures Research Quarterly draft VIII, 1-15, https://pdfs.semanticscholar.org/247d/e073104a6d07537f60b372b711c71935cad6.pdf, (Access date: 03.01.2018)

Ibrahim, F. M., Zubairi Y. Z., Aziz A. A, Zaharim A. (2012) "Using The Rasch Model To Assess Examination Beyond Students' Scores" Global Engineering Education Conference (EDUCON 2012). IEEE Conference Proceedings. doi: 10.1109/EDUCON.2012.6201058

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 (H. U. Journal of Education) 31(2): 346-368, http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/2099-published.pdf, (Access date: 06.09.2020)

Koparan, T., Guven B. (2014). “Determination of the statistical literacy levels towards sampling of 8th grade students.” Elementary Education Online 13 (4): 1171-1184

Linacre J.M. (1989), "Observations are always ordinal; measurements, however, must be interval", Archives of Physical Medicine and Rehabilitation 70(12):857-60

OECD Report. (2012) “Assessment Of Higher Education Learning Outcomes”. AHELO Feasibility Study Report. Vol 1. Design and Implementation.

Öztemel E. (2018). “Eğitimde Yeni Yönelimlerin Değerlendirilmesi ve Eğitim 4.0”, Journal of University Research, April 2018, Volume 1, Issue 1, Page: 25-30, https://dergipark.org.tr/download/article-file/451373, (Access date: 06.12.2018)

Özyurt H., Özyurt O., Baki A. (2012), “Architecture and design process of the ındividualized assesment system ıntegrable to distance education softwares”, Turkish Online Journal of Distance Education-TOJDE July 2012 ISSN 1302-6488 Volume: 13 Number: 3 Article 15

Puncreobutr, V. (2016). “Education 4.0: New Challenge of Learning”, St. Theresa Journal of Humanities and Social Sciences, 2(2), http://www.stic.ac.th/ojs/index.php/sjhs/article/view/Position%20Paper3 (Access date: 03.04.2019)

Waugh, R. F. (1999) “Approaches To Studying For Students In Higher Education: A Rasch Measurement Model Analysis.” British Journal of Educational Psychology 69: 63-79. doi: 10.1348/000709999157572

Wiliam D., (2011). “Embedded Formative Assessment“, p.104. Solution Tree Press, USA

Wiliam D., (2011). “What is assessment for learning?”, Studies in Educational Evaluation 37 (2011) 3–14

Wright B.D., (1977), “Solving Measurement Problems with the Rasch Model”, Journal of Educational Measurement , Vol. 14, No. 2, Applications of Latent Trait Models, pp. 97-116

Wright B.D., Stone M.H. (1979), “Best test design - Rasch Measurement”, Mesa press, Chicago

Yang, SC., Tsou MY., Chen ET., Chan KH., Chang KY. (2011) “Statistical Item Analysis Of The Examination In Anesthesiology For Medical Students Using The Rasch Model.” Journal of the Chinese Medical Association 74: 125-129, doi: 10.1016/j.jcma.2011.01.027