IOJPE

ISSN: 1300 – 915X
www.iojpe.org

International Online Journal of Primary Education

2021, volume 10, issue 1

A RASCH MODEL ANALYSIS OF PRIMARY SCHOOL STUDENTS' CONCEPTUAL UNDERSTANDING LEVELS OF THE CONCEPT OF LIGHT

Hüseyin Cihan BOZDAĞ
Ministry of National Education, Buca Gazi Secondary School, İzmir, Turkey
ORCID: https://orcid.org/0000-0001-6735-7096
chnbzd@gmail.com

Suat TÜRKOĞUZ
Assoc.Prof.Dr. Dokuz Eylul University, Buca Faculty of Education, İzmir, Turkey
ORCID: https://orcid.org/0000-0002-7850-2305
suat.turkoguz@gmail.com

Received: December 05, 2020 **Accepted:** March 08, 2021 **Published:** June 30, 2021

Suggested Citation:
Bozdağ, C. H., & Türkoğuz, S. (2021). A Rasch model analysis of primary school students' conceptual understanding levels of the concept of light. *International Online Journal of Primary Education (IOJPE), 10*(1), 160-179.

Abstract

The study determines the conceptual understanding levels of primary school students on the concept of light according to the Rasch Model with a Four-tier Light Conceptual Understanding Test (LCUT). The participants were 355 (164 girls and 191 boys) primary school students studying at a public school in Izmir city center. In the study, the Rasch Model, which is included in the Latent Trait Theory, was used. Also, the data regarding the answers given and the level of confidence in the responses were associated with the Rasch analysis of LCUT. The results of Rasch analysis showed that LCUT was in full harmony in the context of infit, outfit, and point measurement correlation statistics, and is a valid and reliable measurement tool for conceptual understanding. Moreover, these results explained that the students' average conceptual understanding ability regarding the Light unit was above the average item difficulty.

Keywords: Conceptual understanding, primary school, four-tier diagnostic test, light unit, Rasch model.

INTRODUCTION

The evaluation of the training process is as important as its planning and implementation. Measurement and evaluation make a great contribution to the education process by determining the development levels of individuals' knowledge and skills (Çetin, 2019). However, measurement and evaluation in education is done indirectly with measurement tools consisting of multiple variables, and therefore this contribution is possible with and data based on consistent, valid and reliable measurement processes and measurement tools. Likewise, correct statistics, analysis and calculations may the educational value of these data increase. After all, the sensitivity of the measurement data increases with the quality of structural statistic analyzes used in the measurement tools and contributes to the studies in educational assessment indirectly.

In the measurement and evaluation processes, two theories are used, namely, Classical Test Theory (CTT) and Latent Trait Theory (LTT) (Kan, 2006; Keeves, 1998; Kelecioğlu, 2001). CTT is a long-known theory where evaluation is made according to the whole test, not to each item (Bulut, 2018). Therefore, the contribution of different items to the individual's success is considered to be equal (Anshel, Weatherby, Kang, & Watson, 2009). The search for an alternative theory for CTT has started due to the following limitations: 1) the estimation of individuals' abilities based on their total score, 2) the scores of the individuals depend on the test applied, 3) item statistics depend on the characteristics of the group to which the test was applied (Bulut, 2018; Demirtaşlı, 1996; Hambleton & Jones, 1993; Kelecioğlu, 2001). Accordingly, with the claim that it can overcome these limitations, LTT was introduced at the end of the 1930s (Doğan & Tezbaşaran, 2003). According to LTT, there is a relationship between the individuals' ability, which is not directly observed in a certain area, and their answers to the test items, consisting of questions that examine this area; this relationship can be expressed mathematically (Kelecioğlu, 2001).

The two most widely used and open-to-develop models of LTT developed as an alternative to CTT are the "Item Response Theory" (IRT) and the "Rasch Model" (RM) (Akın & Baştürk, 2012). IRT assumes that a person's performance can be predicted in a test thanks to multiple features (Bulut, 2018). RM, on the other hand, is a technique developed by George Rasch in 1960 and evolved from IRT (Doğru, 2019). RM determines the difficulty levels of the items and the ability levels of the individuals. According to the model, it attempts to determine the probability of what an individual with a certain ability level can do against the desired task (Rasch, 1961).

Although CTT is still widely preferred today in the analysis of a measurement tool, the biggest limitation of CTT is that the characteristics of participants depend on the item and item characteristics on the participant characteristics (Demirtaşlı, 1996). Accordingly, the contribution of different items in predicting individuals' ability levels is equal. On the other hand, RM is different from many other statistical models since it is a probabilistic model (Boone & Scantlebury, 2006). In RM, the characteristics of an item can be calculated independently of the ability level of the participants, and the ability levels of the participants can be estimated independently from the item sample they answered. Also, the contribution of different items in predicting individuals' ability levels is not equal (İlhan & Güler, 2017). According to RM, an individual with a higher ability than others is more likely to correctly answer another item that measures the same structure. Similarly, a question that is easy for any individual is likely to be answered correctly, and a difficult question is unlikely to be answered (Bond & Fox, 2007). Accordingly, the mathematical expression describing the relationship between the test items and the person is shown in Formula 1.

$$P(X_{ni} = 1) = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}}$$

In the formula, (B_n) is a parameter that shows the person's ability. If a person with B_n ability answers a test item with D_i difficulty, it will simply either succeed or fail (Planinic, Boone, Susac & Ivanjek, 2019). Accordingly, the possibility of a person answering an item correctly (P) is related to the person's ability (B_n) and the difficulty of the answered item (D_i) (Boone & Scantlebury, 2006). According to RM formula, the probability of a correct response is expressed as B_n-D_i. If the person's ability equals item difficulty (B_n=D_i), then the probability of a correct response is 50%. If the difference between person ability and item difficulty increases positively (B_n>D_i), then the person has a higher probability of a correct response to the question (Planinic, Ivanjek, & Susac, 2010; Xiao, Han, Koenig, Xiong, & Bao, 2018).

Undoubtedly, in the analysis of a measurement tool with RM, associating the ability of the person with each item (a) facilitates the development of tools that provide useful data, and (b) provides data that can be safely used for both descriptive and parametric statistics (Boone & Noltemeyer, 2017). Therefore, instead of evaluating the raw scores in the analysis of the questionnaires and tests frequently used in education and social sciences, the analyses to be conducted with RM made it possible to reach more objective measurement results that were stripped of many statistical limitations (Boone & Noltemeyer, 2017; Gülkaya, 2018; Preece, 1979; Uzunsakal & Yıldız, 2018). This has enabled RM to have a wide application area and accelerate studies, such as in the fields of health studies, marketing, education, social sciences, and economics.

In studies on education, researchers have started to use RM widely recently (Baharun, Razi, Abidin, Musa, & Mahmud, 2017; Boone & Noltemeyer, 2017; Çetin, 2019; İlhan & Güler, 2017; Maat, 2015; Othman, Salleh, Hussein, & Wahid, 2014). In particular, RM is highly preferred in studies where measurement tools are developed and analyzed in education (Boone & Scantlebury, 2006; Boone, Townsend, & Staver, 2011; Planinic, Ivanjek, & Susac, 2010; Sondergeld & Johnson, 2014; Wei, Liu, Wang, & Wang, 2012). Additionally, conceptual understanding levels can also be analyzed with RM (Liu, 2010; Kauertz & Fischer, 2006; Mešić et al., 2019; Siang, 2011; Wei, Liu, & Jia, 2014).

One of the integral parts and subjects of everyday life in the field of science is light. The light concept is used as the primary tool in many fields from physics to biology, medicine to astronomy. Therefore, students have difficulties in many scientific fields and subjects without understanding the light concept and its properties (Djanette, Fouad, & Djamel, 2013). The studies related to the light concept in the literature focuse on the framework of determining misconceptions (Aydoslu, 2018; Blizak, Chafiqi, & Kendil, 2009; Epik et al., 2002; Fariyani, Rusilowati, & Sugianto, 2017; Galili & Hazan, 2000; Kaplan, 2017; Taşlıdere & Eryılmaz, 2015; Wahyuningsih, Rusilowati, & Hindarto, 2017), determining the conceptual understanding levels within the framework of CTT (Andersson & Bach, 2005; Ayvacı & Candaş, 2018; Demirci & Ahçı, 2016; Kara, Avcı, & Çekbaş, 2008; Şahin, İpek, & Ayas, 2008), mental models (Uzun & Karaman, 2016), cognitive structure (Apaydın, Akman, Taş, & Peker, 2014; Özcan & Tavukçuoğlu, 2018), and teaching methods and techniques (Altun, 2006; Benek & Kocakaya, 2012; Mazlum & Yiğit, 2017; Şenel, 2016). On the other hand, RM-based studies on light are limited (Aminudin, Kaniawati, Suhendi, Samsudin, Coştu, & Adimayuda, 2019). However, the literature has signed that there are not any studies focused on developing a measurement tool on light and analyzing the conceptual level of understanding with RM.

The starting point of this study is the idea that RM provides statistically more reliable evaluation results. The model explains how a person's performance for a particular feature can predict a person's response (eg, true or false) in a particular test item containing that feature (Boone, 2016; Boone & Scantlebury, 2006). These features that are taken into consideration, such as scientific learning, scientific inquiry, or attitude toward science, are defined as "latent/implicit features" (Boone & Scantlebury, 2006; Planinic, Boone, Susac, & Ivanjek, 2019; Xiao et al., 2018). The latent feature used in the assessment of competencies often constitutes a skill. In this study, the latent feature is the conceptual understanding ability about light subject. On the other hand, the quality, reliability, and validity of the test tool used are closely related to the determination of the conceptual understanding level. In this context, the study searches for answers to the questions below.

Research Questions

1- Is the conceptual understanding test developed on the Light unit a valid and reliable measurement tool according to RM?

2- Can RM adequately explain students' conceptual understanding levels on the concept of light?

3- According to RM, what level are the students' conceptual understanding level on the concept of light?

METHOD

Research Model

We analyze LCUT according to the "partial credit or point model" of RM. The model proposed by Masters in 1982 is a suitable model for polytomous items that require multiple stages and are given partial points if different stages are completed during the analysis process (Kaskatı, 2011; Yüksel, 2012). In the Partial Credit Model (PCM), each item has its own ordered scale structure (Gülkaya, 2018). In this model, instead of a binary result as yes/no for items, partial scores can be obtained by considering the answers given when reaching the result (Uzunsakal & Yıldız, 2018). It is useful in situations where students do not mark only as true or false; thus, student competencies can be determined in more detail. So, in the study, we determine the competencies of students on the concept of light from a basic level to a detailed level using PCM.

Participants

The participants were 355 students randomly selected, consisting of 164 girls (46.2%) and 191 boys (53.8%) in the fifth-grade of a state primary school in İzmir city centre. In the study, we chose the sample to represent the general population by using random sampling method from students with similar socio-economic characteristics. The most important feature of this sampling method is that all units in the general population have an equal and independent chance to be selected for the sample (Büyüköztürk, Çakmak, Akgün, Karadeniz, & Demirel, 2014).

Development of Data Collection Tool

In the study, we developed the Four-tier Light Conceptual Understanding Test (LCUT) and analyze it with RM. We used a 4D model (Defining, Designing, Developing, Disseminating), which helps the researcher to design a product that will help the student develop their skills in the learning process (Irawan, Nyoman Padmadewi, & Artini, 2018), in the development of LCUT. In the 'defining' stage, we conducted a literature review on the subject of light and four-tier tests. During the 'designing' stage, we examined the structure of the four-tier test. The first tier of the four-tier test is the multiple choice question tier; the third tier is the reasoning tier for the response to the first tier. The second and the fourth tiers are the confidence tiers. There are six options which are rated between "1" and "6" in the confidence tier, respectively: "Just guess", "I'm not too sure", "I'm not sure", "I'm sure", "I'm pretty sure", and "I'm absolutely sure." The design of the four-tier test is shown in Figure 1.

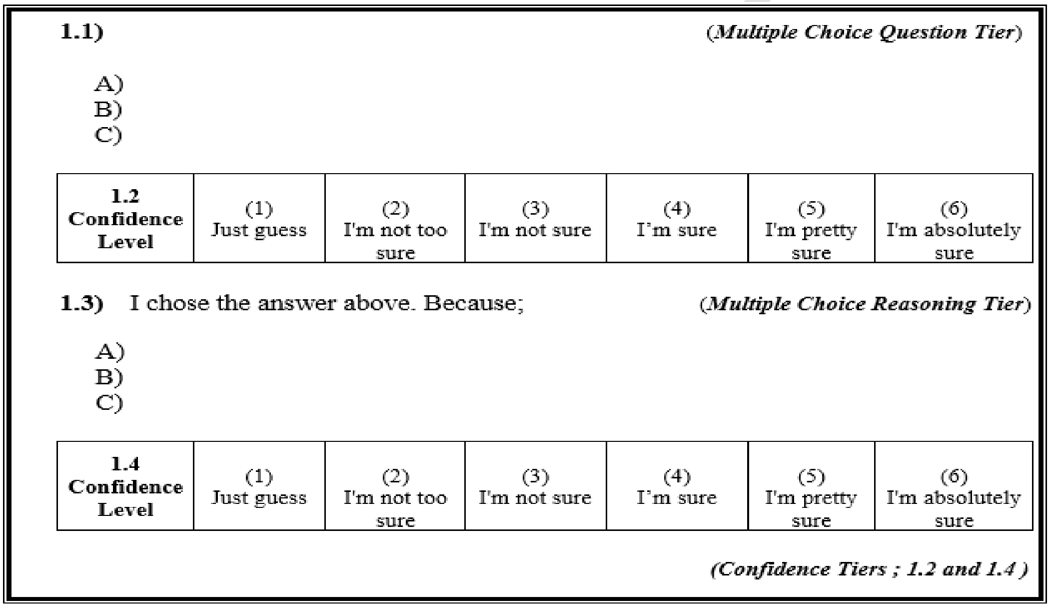

Figure 1. The design of Four-tier test

In the 'developing' stage, we considered studies on the literature about conceptual understanding levels and misconceptions on the subject of light, achievements regarding the subject of light in the curriculum of the science course, textbook, achievement tests, the suggestions of the course teachers and experts, and the opinions of the students and the answers given in the open-ended exams. In this stage, we developed LCUT consisting of fourteen questions and then we examined these items by the opinions of a faculty member, a field expert, and two science teachers. Considering experts' evaluations, we took into two questions determined to have the same content into the same question root, and canceled one question. A Turkish teacher and 20 students in higher education examined LCUT, which consisted of twelve questions, in terms of reading and understanding. In accordance with the feedback of languege expert, we eliminated the deficiencies in the form of spelling errors and we developed the final measurement tool. We applied LCUT to 355 fifth-grade students in approximately one class hour. Sample questions related to the questions in LCUT are presented in Figure 2. The distribution of the questions in LCUT according to the achievements of the science education program (MNE, 2018) is given in Table 1.

IOJPE
ISSN: 1300 – 915X
www.iojpe.org
International Online Journal of Primary Education
2021, volume 10, issue 1

Table 1. Achievement distribution of questions

UNIT	SUBJECT	ACHIEVEMENT	QUESTION
SPREADING THE LIGHT	Spreading the light	1-Observing that the light coming from a source follows a linear path in all directions, it shows with drawing.	Q1, Q10
	Reflaction of light	1-Observes the reflactions of light on smooth and rough surfaces and shows them by drawing.	Q4, Q6
		2-Explains the relationship between the incoming beam, reflected beam and the normal of surface.	Q2, Q3
	Encounter of light with a substance	1-Categorizes the substances according to their light transmittance status.	Q7, Q8
	Full shadow	1-Observes how the full shadow is formed and shows it with simple beam drawings.	Q11, Q12
		2-Discovers the variables that affect the full shadow with experiments.	Q5, Q9, Q12

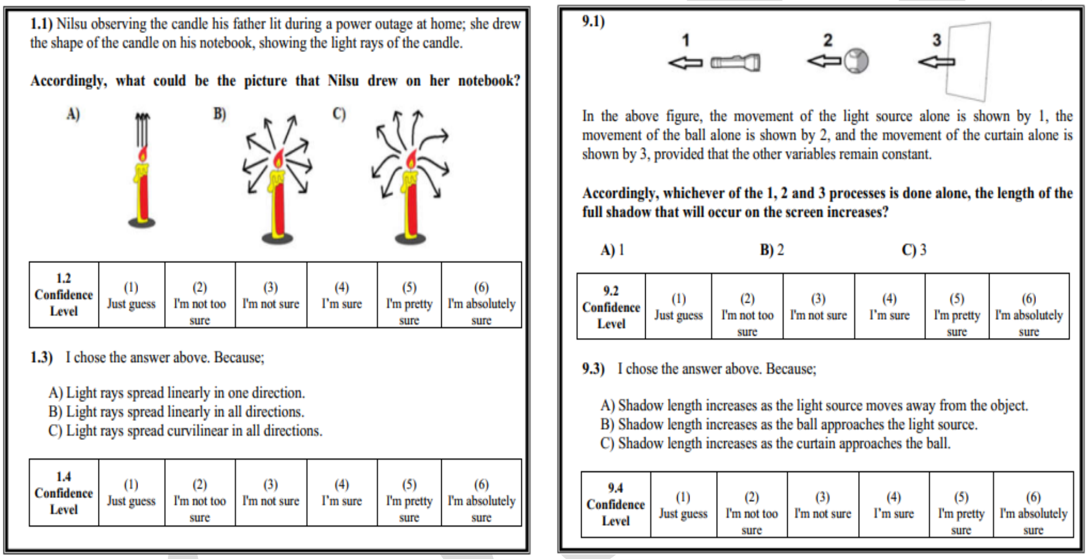

Figure 2. LCUT sample questios

Analysis of Data

Scoring Categories for LCUT

Due to the four-tier test, we gathered the answer combinations for the responses given to the tiers in each test item under six categories as Scientific Knowledge, Misconception, Lucky Guess, False Positive, False Negative, and Lack of Knowledge (See Table 2).

Table 2. Scoring Categories for LCUT

CATEGORY	SK	*LG*			FP	FN	MC	*LK*								
1.TIER	T	*T*	*T*	*T*	T	F	F	*T*	*T*	*T*	*F*	*F*	*F*	*F*	*F*	*F*
2.TIER	S	*S*	*NS*	*NS*	S	S	S	*S*	*NS*	*NS*	*S*	*S*	*NS*	*NS*	*NS*	*NS*
3.TIER	T	*T*	*T*	*T*	F	T	F	*F*	*F*	*F*	*T*	*F*	*T*	*T*	*F*	*F*
4.TIER	S	*NS*	*S*	*NS*	S	S	S	*NS*	*S*	*NS*	*NS*	*NS*	*S*	*NS*	*S*	*NS*

SK: Scientific Knowledge, LG: Lucky Guess, FP: False Positive, FN: False Negative, MC: Misconception, LK: Lack of Knowledge, T: True, F: False, S: Sure (Confident Level>3,5), NS: Not Sure: (Confident Level<3,5)

Although six scoring categories are used for four-tier tests, in the study, we determined the conceptual understanding levels of students based on a Rasch analysis, and therefore, we arranged the 3 different scoring categories for the correct answers given at each tier for RM: Conceptual Understanding, Misconception, and Confidence Level (See Table 3).

Table 3. Scoring categories for Rasch analysis

CATEGORY	TIER	SCORE	EXPLANATION
CONCEPTUAL UNDERSTANDING	1.Tier	1	If the student responds correctly to 1. Tier (Question Tier)
		0	If the student responds incorrectly to 1.Tier
	3.Tier	1	If the student responds correctly to 3. Tier (Reasoning Tier)
		0	If the student responds incorrectly to 3.Tier
	1. and 3. Tier	1	If the student responds correctly to 1.and 3. Tier
		0	In all other alternatives
	All Tiers	1	When the student responds correctly to 1. and 3.Tier with confidence
		0	In all other alternatives
MISCONCEPTION	1.Tier	1	If the student responds incorrectly (accompanied by misconception) to 1. Tier
		0	If the student responds correctly to 1.Tier
	3.Tier	1	If the student responds incorrectly (accompanied by misconception) to 3. Tier
		0	If the student responds correctly to 3.Tier
	1. and 3. Tier	1	If the student responds incorrectly to 1. and 3. Tier
		0	In all other alternatives
	All Tiers	1	If the student responds incorrectly to 1. and 3.Tier with confidence
		0	In all other alternatives
CONFIDENT LEVEL	2.Tier	1	CL >3.5 (Sure)
		0	CL < 3.5 (Not Sure)
	4.Tier	1	CL >3.5 (Sure)
		0	CL < 3.5 (Not Sure)
	2. and 4. Tier	1	CL >3.5 (Sure) for 2, and 4. Tier
		0	In all other alternatives

CL: Confident Level, The threshold value was considered as 3,5 to determine CL

Table 3 shows that since a certain score is taken based on the responses given in each tier of each test item, PCM was used in the analyses to be conducted. Within the framework of PCM, partial scores can be obtained based on the responses given for the tiers of each item of the four-tier test. Therefore, we converted response alternatives related to the tiers in each test item specified in Table 2 were converted into partial scoring in line with the scoring categories in Table 3.

Table 4. PCM scoring key

CATEGORY	SCORES		
	Conceptual Understanding	Misconception	Confident Level
SK	4	0	
LG	3	0	
FP	1	1	
FN	1	1	
LK (Alternatives with correct answers)	1	1	
LK (Alternatives with wrong answers)	0	3	
MC	0	4	
S			3
PS			1
NS			0

SK: Scientific Knowledge, LG: Lucky Guess, FP: False Positive, FN: False Negative, MC: A Misconception, LK: Lack of Knowledge, S: Sure (Confident Level [CL]>3.5 for 2-4.tiers), NS: Not Sure: (CL<3.5 for 2-4 tiers), PS: Partial Sure (CL>3.5 for 2.tier, CL<3.5 for 4.tier or vice versa)

As a result, the scoring key (0-1-3-4) for conceptual understanding, misconception, and confidence level was created for the four-tiers (See Table 4).

Rasch Model Analysis Findings

We analysed by the data through WINSTEPS software to answer research questions. In the first part, we tested LCUT's reliability and its validity. In the second part, we examined students' responses on LCUT in the scope of conceptual understanding, misconception, and trust level.

Findings on Validity and Reliability

In this part, we evaluated LCUT's reliability and its validity through WINSTEPS software. In Rasch analysis, reliability is examined under two headings: person reliability and item reliability. Person reliability refers to the consistency of student responses and item reliability refers to the quality of test items. Accordingly, Table 5 shows the analysis results on person reliability for students and item reliability for LCUT's items.

Table 5. Rasch analysis results about person and item for LCUT

PERSON	Total Score	Count	Measure	Model S.E.	Infit Mnsq	Infit ZSTD	Outfit Mnsq	Outfit ZSTD
MEAN	28.8	12.0	52.48	2.23	1.01	.0	1.02	.1
P. SD	9.9	.0	4.41	.65	.38	1.0	.68	.8
S. SD	10.0	.0	4.42	.65	.38	1.0	.68	.8
MAX.	47.0	12.0	66.39	7.48	2.56	2.5	5.67	3.1
MIN.	8.0	12.0	43.06	1.90	.18	-2.9	.17	-1.9
REAL RMSE=2.51, TRUE SD=3.63, SEPARATION=1.45, Person RELIABILITY=.68, S.E.of Person Mean=.24								
MODEL RMSE=2.53, TRUE SD=3.75, SEPARATION=1.61, Person RELIABILITY=.72								
ITEM	**Total Score**	**Count**	**Measure**	**Model S.E.**	**Infit Mnsq**	**Infit ZSTD**	**Outfit Mnsq**	**Outfit ZSTD**
MEAN	866.1	355.0	50.0	.41	.98	-0.1	1.02	.4
P. SD	264.3	.0	4.10	.08	.13	1.7	.26	1.7
S. SD	276.1	.0	4.28	.08	.13	1.7	.27	1.7
MAX.	1300.0	355.0	55.94	.60	1.19	2.7	1.39	2.9
MIN.	456.0	355.0	42.33	.36	.81	-3.1	.54	-1.9
REAL RMSE=.42, TRUE SD=4.08, SEPARATION=9.62, Item RELIABILITY=.99, S.E. of Item Mean = 1.24								
MODEL RMSE=.42, TRUE SD=4.08, SEPARATION=9.80, Item RELIABILITY=.99								

The results in Table 5 show that while the person reliability for LCUT is in the range of .68-.72, the Cronbach alpha value is .76 and the item reliability for LCUT is .99. In this case, are the results reliable or not? To decide to this, the ideal value of person reliability should be greater than .80 (Bond & Fox, 2007; Linacre, 2014), however, the values greater than .60 can be accepted as reliable, repeatable, and valid for measurement (Zain, Mohd, & El-Qawasmeh, 2011), and even values greater than .67 can be considered reasonable (Fisher (2007). Moreover, person reliability is also equivalent to Cronbach alpha (KR-20), which is CTT reliability (Linacre, 2014). For a measurement tool to be considered reliable, the Cronbach alpha coefficient must be greater than .70 (Büyüköztürk et al., 2014). In additon, item reliability, the acceptable value should be greater than .80, and values in the range of .67-.80 may be reasonable (Fisher, 2007; Linacre, 2014). Overall, the results in Table 5 show that the person, Cronbach's alpha and item reliability coefficients for the measurement reliability of LCUT can be accepted within the range of the reliability criteria.

A Rasch analysis allows person and item separation coefficients to evaluate internal consistency of a test. The person separation coefficient refers to the range of measured ability scores, and the item separation coefficient refers to the spread of item difficulty levels. Separation coefficient values are between 0 and infinite, and higher values indicate better separation (Boone & Noltemeyer, 2017). Linacre (2019) states that the item separation coefficient should be 3 or above. In terms of person separation coefficient, 1.50 is weak, but acceptable, 2.00 is good, and 3.00 is considered excellent (Duncan et al., 2003; Fisher, 2007; Linacre, 2019). Therefore, Table 5 shows that both the person (1.61) and item (9.80) separation coefficients of RM are appropriate and acceptable for the test's internal consistency.

The determining that the scores of a measurement tool are meaningful, useful, and purposeful by analyzing the construct validity results in RM is likely. The point measurement correlation (Ptmea Corr), infit mnsq and outfit mnsq statistics in the analysis of RM can determine the construct validity of a test. Table 6 shows that the results of the Ptmea-Corr analysis on each item in LCUT.

Table 6. Item analysis results based on RM for LCUT's construct validity

ITEM	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E	INFIT		OUTFIT		PTMEA-AL		EXACT MATCH	
					MNSQ	ZSTD	MNSQ	ZSTD	Corr.	Exp.	OB%	EXP%
S1	1300	355	42.33	.60	.82	-1.2	.54	-1.9	.36	.27	76.6	73.2
S2	664	355	53.00	.36	1.06	.9	1.09	.9	.52	.55	23.4	23.7
S3	1068	355	47.48	.40	.87	-1.8	.83	-1.1	.49	.43	42.2	37.4
S4	1006	355	48.43	.38	.81	-3.1	.76	-1.9	.53	.46	40.8	33.8
S5	456	355	55.94	.39	1.11	1.5	1.39	2.8	.50	.57	24.3	32.1
S6	1269	355	43.35	.54	.82	-1.5	.69	-1.3	.39	.30	70.5	65.0
S7	782	355	51.46	.36	1.15	2.3	1.16	1.6	.48	.53	21.7	21.1
S8	840	355	50.70	.36	1.02	.4	1.02	.2	.52	.51	19.9	20.6
S9	776	355	51.54	.36	1.02	.3	1.04	.4	.51	.53	22.3	20.6
S10	1058	355	47.64	.40	.95	-.6	1.36	2.2	.43	.44	36.1	35.1
S11	657	355	53.09	.36	.96	-.6	.98	-.1	.56	.55	21.7	23.7
S12	517	355	55.02	.38	1.19	2.7	1.36	2.9	.48	.56	18.5	27.2
Mean	866.1	355	50.00	.41	.98	-.1	1.02	.4			34.8	34.4
P.SD	264.3		4.10	.08	.13	1.7	.26	1.7			19.0	16.6

In Table 6, the Ptmea-Corr values are in the range of .36-.56 for LCUT's items. While Bond and Fox (2007) stated that the correlation coefficient should have values greater than .30, Othman et al. (2014), on the other hand, stated that values less than .35 mean weak and low correlation, values between .36-.67 mean moderate and reasonable correlations, values between .68-1.00 represent strong and high correlation values. The results in Table 6 explain that the Ptmea-Corr values have a moderate and reasonable correlation in comparison with these threshold values (Bond & Fox, 2007). This means that each test item can distinguish the ability of the participants.

Fit statistics in RM were also evaluated for determining the construct validity. Linacre (2002) states that fit statistics values between .50-1.50 are suitable values for measurement. Considering these statistics in Table 6, the values of infit mnsq statistics are in the range of .82-1.19 and outfit mnsq statistics in the range of .54-1.39. In addition, the standardized z values (ZSTD) in Table 6 are between -3.1/2.7 for infit, and are between -1.9/2.9 for outfit. ZSTD values should be between -2.0/2.0, but ZSTD values that are not within the desired limits can be ignored if infit and outfit mnsq statistical values have acceptable values (Bond & Fox, 2007; Linacre, 2014). Therefore, these results indicate that all the items in LCUT are in harmony for the measurement in the range of Linecra's fit statistics values.

RM is based on the unidimensionality of the test items. In other words, the test is expected to measure a single structure. Linacre (2006) suggests to check the multidimensional of test. In this case, Principal Compenent Analysis (PCA) should be performed to evaluate dimensionality. So, the data obtained from the Winsteps software were applied PCA. Consequently, the raw variance explained by measures for LCUT was 42.3% and, the eigenvalues of the unexplained variance in 1st, 2nd, 3rd, and 4th contrasts were 1.4, 1.3, 1.2, and 1.1, respectively. The results verify that data have a unidimensionality within the criteria determined by Linacre (2006). Overall, we can suggest that the conceptual understanding test (LCUT) developed on the concept of light is a valid and reliable measurement tool according to RM.

Findings on Person-Item Wright Map

RM allows the comparison of difficulty levels of each test item with the person's ability level in a common metric unit of measurement. Therefore, the Winsteps software allowed the association of item difficulty level with person ability levels by converting them into common measurement values. Thus, a common measurement unit on the same linear scale called Person-Item Map (Wright Maps) can explain person ability and item difficulty measurements (Bond & Fox, 2007). Accordingly, the person-item map in Figure 3 shows the distribution of test items and persons in a common measurement metric. The person-item map is arranged in two vertical histograms. The right side is for participating students, and the left side is for test items. Vertical movement on the person-item map indicates that the ability of the participants ranked from highest to lowest in the right side, and test items ranked from hardest to easiest in the left side. “M”, on the separation line in the middle of the person-item map, means the average of item difficulty and person ability, “S” (Single) means one standard deviation distance from the average, “T” (Twice) means two standard deviations from the average.

```
MEASURE      Item - MAP - Person
              <rare>|<more>
 67             + 0160 0196 0207 0248 0288 0337 042  065  072
 66             + 0170 0218 0341
 65             +
 64             +
 63             + 0141 0241
 62             +
 61            +T 0140 0142 0200 0321 0329
 60             + 0126 0147 0162 0176 0201 0203 0213 0220 0242 0249
                  0253 0286 0287 0316 0336 035  039  041  073
 59             + 0155 0177
 58            T+ 0179 0223 0225 0298 0309 0317 049  059  086
 57            +S 0123 0129 0137 0150 0153 0169 0174 0197 0204 0210
                  0212 0215 0219 0238 0270 0296 0310 0319 0322 0323
                  0330 0331 044  048  050  071  078  088  090
 56        S5   + 0105 0116 0119 0136 0180 0183 0189 0194 0265 0280
                  0302 0342 0343 0349 06   094
 55       S12   + 0100 0122 0131 0134 0143 0144 0159 0161 0198 0208
                  0217 0226 0234 0244 0251 0261 0264 0281 0292 03
                  030  0305 031  0312 0314 0346 0348 037  045  060
                  07   070  095
 54            S+ 01   012  0139 0172 0205 0221 0227 0228 023  0243
                  0262 0268 0324 036  067  076
 53    S11 S2   + 0107 0111 0125 0151 0152 0190 0202 0224 0233 0237
                  0254 0260 0297 0308 0311 0338 0339 0350 051  055
                  057  085  09   098
 52         S9 +M 0101 0103 0104 0109 0112 0128 013  0138 0145 0148
                  0158 0163 018  0193 0195 020  0214 025  0252 0258
                  026  0266 0283 0289 0294 0300 0301 0306 0318 0320
                  0355 04   046  053  062  064  079  082  091  092
                  097
 51     S7 S8   + 0132 0133 0135 014  0188 0211 0222 0232 0236 024
                  0271 0303 0335 0353 047
 50            M+ 0106 0114 0146 0149 0156 016  0164 0168 0175 0186
                  0187 0192 021  0216 022  0231 0235 0255 0256 0279
                  028  0304 033  0332 034  0340 0344 0345 0354 040
                  058  066  083  087  089  096
 49             + 0102 0108 0118 0120 0121 0165 0167 0181 0185 019
                  0199 0209 0229 0239 0245 0246 0267 0269 0272 0273
                  0290 0291 0307 0313 0315 032  0327 0328 0347 0351
                  052  054  069  077  08   099
 48    S4 S10  +S 0113 0115 0117 0127 0154 0166 0184 02   0278 043
                  068  074
 47        S3   + 011  0110 0130 015  017  0240 0257 0282 0284 029
                  0299 0333 05   061  080  081  084  093
 46            S+ 010  0124 0157 0171 0247 0259 0263 0275 0276 0295
                  063  075
 45             + 0173 0191 0206 0230 0250 027  0274 0285 0293 0352
 44            +T 0178 0182 0277 0326 0334 038  056
 43        S6   + 0325
 42        S1  T+
              <freq>|<less>
```

Figure 3. Conceptual understanding person-item map

Figure 3 shows a great distribution of the test items according to difficulty level and the person ability level. This distribution indicates that the item and person separation are high and test difficulty is appropriate. On the other hand, since the Winsteps software determines a common measurement value, it makes possible to compare the average values of person ability and item difficulty.

Accordingly, the average measurement values in the distribution is 50 for item difficulty and is 52 for person ability. Consequently, these values are quite close to each other. This indicates that the range of test items is suitable for the group of participants, that is, test items are not too difficult or too easy for students. According to the person-item map, the questions below the average item difficulty (S1-S3-S4-S6-S10) are relatively easy questions, while the questions above the average item difficulty (S2-S5-S7-S8-S9-S11-S12) are relatively difficult questions. On the basis of person's ability, we determined that the questions equal to and above the average person ability were the items coded as S2, S5, S9, S11, S12 (difficult for students), and also the most difficult questions were S5 and S12. On the other hand, although items coded as S7 and S8 are above the average item difficulty, they remain below the average person's ability. Accordingly, items coded as S1, S3, S4, S6, S7, S8 and S10, which are below the average person ability, are easy questions for students. The questions coded as S1 and S6, which all participants responded correctly, are below the lowest person ability.

Undoubtedly, this assessment is likely to conduct in the context of average person ability and item difficulty measurement for each person. Therefore, these findings of Rasch analysis indicate that RM allows to determine the conceptual understanding level regarding the subject of light.

Findings on Students' Conceptual Understanding Levels

The person-item maps can help to identify students with the highest ability and lowest ability. For example, Figure 3 shows that the ability level of many students is even above the item coded as S5, which is determined to be the most difficult question. Therefore, the ability level of students who are above the item coded as S5 in Figure 5 is above the difficulty level of all questions and is at a level that can answer all questions correctly. Among the students who are above the item coded as S5, a total of nine students coded as O42, O65, O72, O160, O196, O207, O248, O288, and O337, have the highest conceptual understanding level. On the other hand, the student coded as O325, which is at the bottom of the person-item map, has the lowest ability.

In Figure 3, the number of students at and above average person ability is 207. Accordingly, 58.3% of the students (207/355*100=58.3) are at or above the level of average person ability. Considering this result, the conceptual understanding level of students about the concept of light is at a medium level. While the questions above the average person ability (S2-S5-S9-S11-S12) are relatively difficult for students, the questions below the average person ability (S1-S3-S4-S6-S7-S8-S10) are easy questions.

In the evaluation of the achievement distribution of the questions, the students have low conceptual understanding abilities in terms of full shadow (S5-S9-S11-S12) and reflaction laws (S2), while they have high conceptual understanding abilities in light diffusion (S1-S10) and the reflaction of light (S3-S4-S6). In addition, the conceptual understanding abilities in the subject of light with a substance (S7-S8) are relatively moderate because the fact that question items coded as S7 and S8 are below the average ability level.

RM also provides a comparison of students' ability with other variables. Figure 4 shows that comparison of students' ability levels with confidence levels. In the Figure 4, the dashed blue lines show the relationship between the students, while the dashed red lines indicate the relationship between the questions.

Figure 4 explains that 275 students (77.5%) answered all questions with a high level of confidence. According to the conceptual understanding person-item map, 58.3% of the students were evaluated in the scientific knowledge category by answering the questions correctly with a high level of confidence. This means that there are students who think that the wrong answers are correct within the group of students with a high level of 77.5% confidence. Students in this group are evaluated in the category of misconception due to their wrong answers.

When Figure 4 was examined in the context of questions, items coded as S1 and S6 are at the lowest difficulty level and were responded highly confidently by all students. In fact, there is a similar situation in questions coded as S3, S4, and S10. Students responded to items coded as S3, S4, and

IOJPE

ISSN: 1300 – 915X
www.iojpe.org

International Online Journal of Primary Education 2021, volume 10, issue 1

S10, which were not difficulty to answer, with a high level of confidence. On the other hand, although items coded as S5, S9, S11, and S12 were difficult questions, many students responded these items with a high level of confidence. Responding by students with high confidence in difficult items as well as in easy items indicates that the level of conceptual understanding in these questions is low, but the possibility of misconceptions is high.

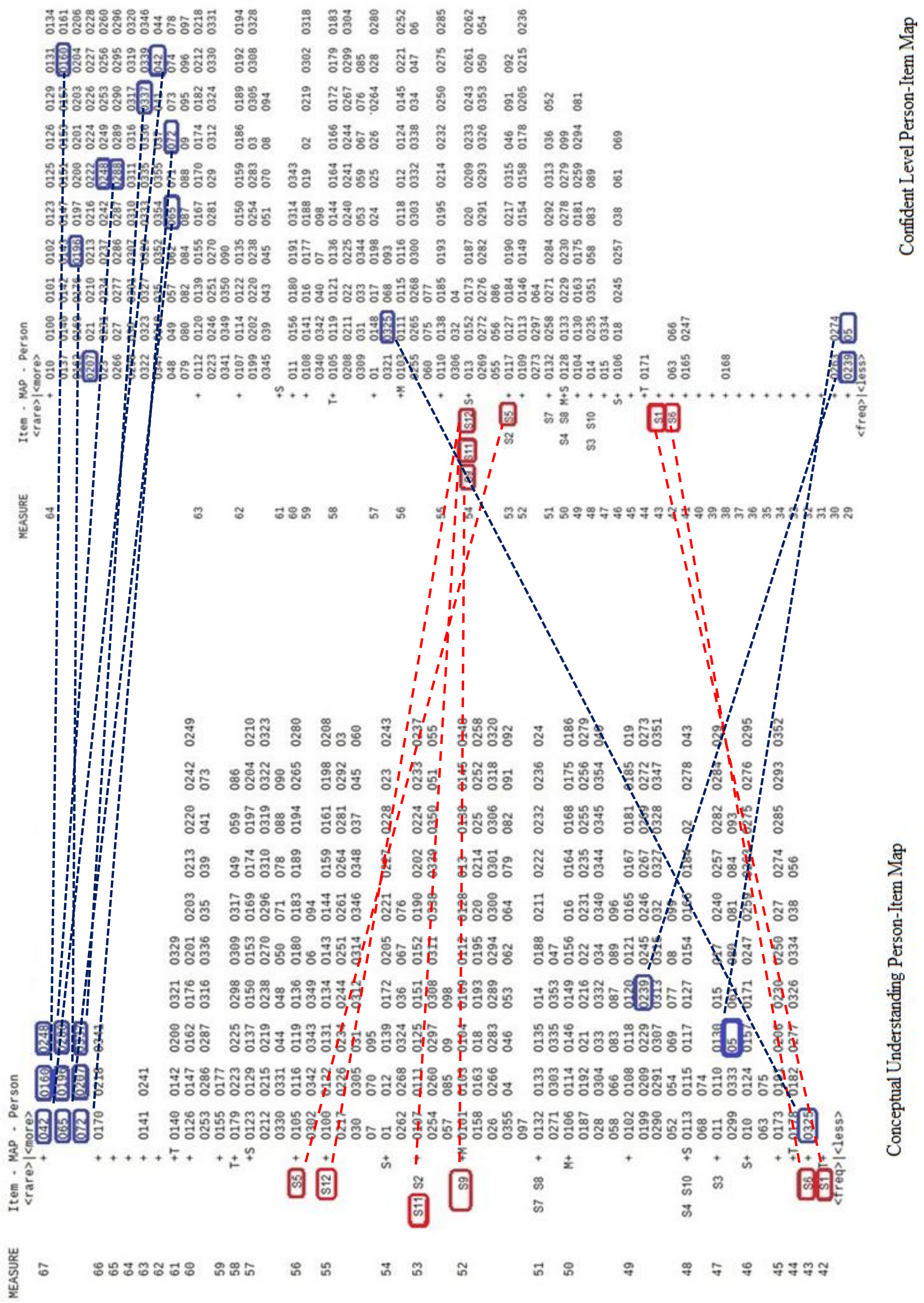

Figure 4. Comparison of conceptual understanding and confidence level person-item maps

IOJPE
International Online Journal of Primary Education
ISSN: 1300 – 915X
www.iojpe.org
2021, volume 10, issue 1

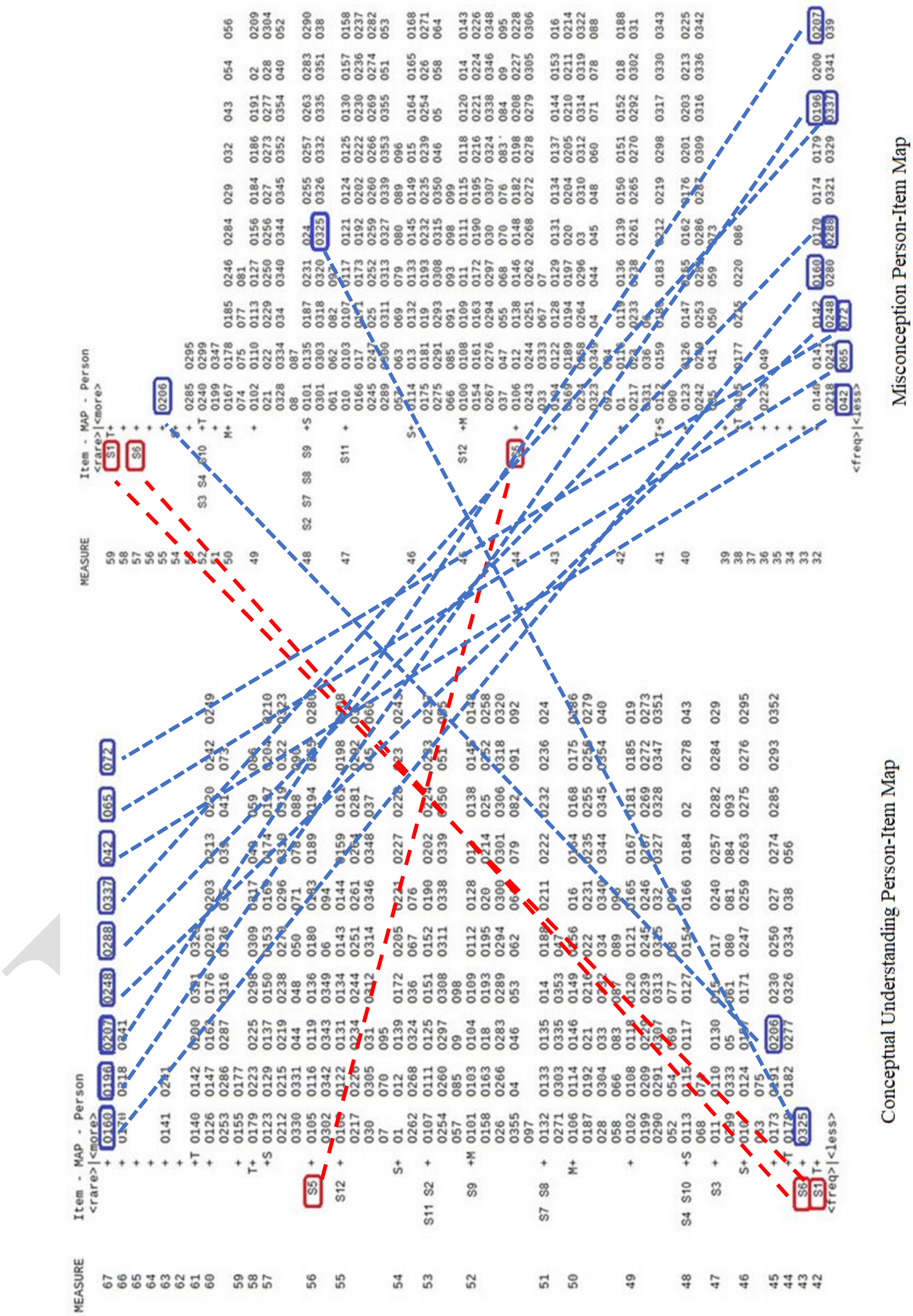

Figure 5. Comparison of Conceptual Understanding and Misconception Person-Item Maps

As a result of the comparison of conceptual understanding and confidence level for person ability in Figure 4, students coded as O42, O65, O72, O160, O196, O207, O248, O288, and O337 have the highest person ability and confidence level. This high connection shows that students are at a high cognitive level in terms of conceptual understanding, that scientific knowledge is fully and accurately placed in their cognitive structures, and therefore they are very confident in their responses. On the other hand, student coded as O325 has the lowest ability level and a medium level of confidence in terms of conceptual understanding. This low connection indicates that student coded as O325 does not reach correct answers in terms of conceptual understanding, but is confident in most of the answers to the questions in the test.

As a result of the comparison of conceptual understanding and confidence level's person item maps, achieving the result that students are sure of their answers, even if they respond incorrectly, required a comparison on the misconception person-item map in the continuation of this study.

Figure 5 provides a comparison of the conceptual understanding person-item map and the misconception person-item map. In the evaluation map of the conceptual understanding, the questions that are difficult to answer by the students are at the left top of the map, the questions that are easy to answer are at the left bottom. Moreover, in the misconception person-item map, the questions that are answered incorrectly (the questions with the misconception) are below in right of the map.

According to Figure 5, the question, which is the most difficult to answer and best determines the misconception, is the test item coded as S5. In other words, the students answered the item coded as S5 with the highest rate of misconception. Additionally, the questions that are difficult to answer correctly and that best determine the misconception are the test items coded as S11 and S12 after the question item coded as S5. On the other hand, the questions, which are the easiest questions under all students' abilities and unable to detect the misconception, are the test items coded as S1 and S6. These data coincide with the data obtained from the person-item map analysis.

In Figure 5, students coded as O42, O65, O72, O160, O196, O207, O248, O288, and O337 have high ability in terms of conceptual understanding and have the lowest potential in terms of misconception. Similarly, the findings show that student coded as O325 has the lowest conceptual understanding ability and the highest potential misconceptions, and different student coded as O206 has the highest misconception potential and the lowest conceptual understanding.

DISCUSSION and CONCLUSION

In this study, we preferred RM, which provides a detailed analysis of the data obtained from the measurement tools and allows the researchers to convey the test-scale performance in the best way (Boone & Noltemeyer, 2017). In this context, we developed a four-tier LCUT with RM and then analyzed the data obtained from this scale. In the development process of LCUT, we provided LCUT's validity and reliability with RM.

In terms of validity, RM analysis provides appropriate data on the construct validity of the measurement tool (Wei, Liu, Wang, & Wang, 2012). Accordingly, we evaluated infit and outfit mnsq statistics to determine the construct validity. All these statistics were in the range of .54-1.39. These results show that all the items in LCUT are in the appropriate range for measurement (Linacre, 2002), and are in harmony. Additionally, we analyzed Ptmea Corr statistics for the consistency between the scores of the students in the test items and their ability measurements. As a result of the analysis, we determined that all items were in the desired range (in the range of .36-.56) and positive. According to the infit and outfit mnsq statistics, all items fit RM very well. As a matter of fact, since Rasch measurement is based on individual response models that reflect individuals' reasoning skills, good model-good data compliance shows that students' reasoning about the measurement tool items is consistent (Wei, Liu, Wang, & Wang, 2012).

When the data fit RM, the test items measure the intended unidimensional structure (Wei, Liu, Wang, & Wang, 2012). RM is a unidimensional measurement model (Eggert & Bögeholz, 2010; Linacre, 2006; Planinic, Ivanjek, & Susac, 2010); unidimensionality is the basic condition of construct validity (Rasch, 1961). In this context, we conducted a unidimensional assessment with the PCA for determining the construct validity of LCUT. Since the results obtained for the explained variance and the unexplained variance are within the specified criteria, we determined that the data showed a unidimensional feature.

To determine the measurement reliability of LCUT, we examined person and item reliability, and also separation coefficient values. First, we determined that the person reliability was .68-.72. We interpreted this reliability in the same way as traditional Cronbach Alpha (Linacre, 2014), and concluded that it was reliable, repeatable, and reasonable value in the .67-.69 measurement range (Fisher, 2007; Zain, Mohd ve El-Qawasmeh, 2011), and therefore the person reliability of IKAT was satisfactory level. In fact, considering CTT perspective, this reliability value may indicate the existence of a strong relationship between the observed person scores and the true scores without errors (Cronbach alpha=.76). Second, we found that item reliability for the quality of test items was .99. In the literature, values of .80 and above are accepted appropriate for item reliability (Fisher, 2007; Linacre, 2014), and therefore we decided that the item reliability of LCUT was appropriate value. The separation coefficient is the spread estimate of items and ability, since they are expressing statistically different levels of performance-difficulty. In the study, we observed that the separation coefficient was 1.61 for the person and 9.80 for the item. In addition, according to Linacre's (2016) guidelines, it is possible to make an assessment by associating the reliability of the person with the level of performance that the test can distinguish (separation unit 3-4 for .90, separation unit 2-3 for .80 and separation unit 1-2 for .50). The person separation coefficient determined as 1.61 indicates that people can be evaluated at two performance levels. Therefore, we can talk about the existence of a match between the person reliability coefficient (.68-.72) and the determined performance levels. Consequently, the fact that the person and item spread estimations specified as the separation coefficient are within the specified criteria also contributes to the reliability of the measurement tool (Duncan et al., 2003; Fisher, 2007; Linacre, 2019).

Rasch analysis has many advantages in determining students' competencies. The most important advantages are the representation of the person and test items on the same equal linear scale (Bond & Fox, 2007). In this way, RM enables the comparison of the ability level of the participants and difficulty level of the test items by matching them to the determined theoretical latent feature with positioning in a common metric (Clements, Sarama, & Liu, 2008). The common metric where the ability level and test items are located is defined as the "Wright Map." In the conceptual understanding person-item map, it was determined that 207 of 355 students were above the average ability level, and that the conceptual understanding level was 58.3%. Therefore, on the map, we observed that the ability level of more than half of the students is above the difficulty level of the questions. Rasch analysis can help to determine the conceptual understanding level with item difficulty level in its model (Wei, Liu & Jia, 2013). Considering this opinion, we found that the students' ability is above the average item difficulty level, that the majority of the students match the items difficulty level very well. In other words, the vast majority of the students have the ability to understand the latent feature to be measured (Baharun et al., 2017). On the other hand, it is possible to find the study findings that the difficulty of the item is higher than the person's ability, and therefore the participants have low conceptual understanding ability (Siang, 2011). In limited studies conducted with Rasch analysis on the concept of light (Aminudin et al., 2019; Mesic et al., 2019), results indicated that the average difficulty level was above the average ability, and the students were at a low conceptual understanding level. Therefore, the study results that were based on the analysis of the four-tier test with RM are important in terms of contributing to the limited study findings in the literature.

The literature emphasizes the need to conduct a confidence level analysis to see to what extent students believe in their abilities (Aminudin et al., 2019). In the confidence level analysis of the study, students answered the questions with a 77.5% confidence level throughout the test. In this context, the study results indicate that students' confidence levels in their answers were quite high. On the other hand, the clear difference between the proportion of students who are evaluated at the scientific knowledge level by responding correctly to all tiers (58.3%) and the confidence level in the test (77.5%) indicates that some students are quite sure of these answers, even if they responded incorrectly. This difference points out the existence of students who are evaluated in the misconception category as well as highly skilled students among high-confidence students. In a comparison of the conceptual understanding person-item map with the misconception person-item map, we determined that the data on the maps confirm each other, and that there was a linear relationship between the students' confidence level and their misconceptions. Indeed, Aminudin et al. (2019) determined that high-ability students are at a low level of misconception and low ability students at a high level of misconception as a result of their studies.

In the comparison of the conceptual understanding level and the difficulty levels of the items, we determined that the students had difficulty in answering some questions of the test (S5-S12), so the difficulty level of these questions was above the ability level of most students. On the other hand, all students answered two questions (S1-S6) of the test, correctly. In other words, the ability level of all students are above the difficulty level of these questions. As a matter of fact, according to the Rasch analysis, the person is more likely to respond to a lower item on the scale and less likely to respond to a higher item on the scale (Boone, 2016). On the other hand, it is possible to determine the students with the lowest and highest conceptual understanding on person-item maps (Bond & Fox, 2007). In this aspect, we determined that the student with the lowest ability in terms of conceptual understanding was student coded as O325, and the students with the highest ability were students coded as O42, O65, O72, O160, O196, O207, O248, O288, and O337. Wright maps allow item difficulty and person ability to be compared (Liu, 2010). As a matter of fact, thanks to the person-item maps showing the data of both participants and test items, the most difficult and easiest questions of the test can be easily determined without the need for another calculation.

In the person-item map of the conceptual understanding, we determined that the seven questions that were above the average item difficulty were the questions that the students had difficulty in answering. We listed these questions in the turn of items coded as S5, S12, S2, S11, S9, S7, and S8 starting from the most difficult one. Items coded as S1, S3, S4, S6, and S10 that are below the average item difficulty level, are the questions that the students have no difficulty in answering. Four questions (S5-S9-S11-S12) in LCUT are to measure students' the conceptual understanding of full shadows. These questions have difficult to answer, therefore, students' conceptual understanding is the lowest level in this test items. As a matter of fact, studies on the concept of light indicate that students have lacks of knowledge about shadow formation and shape (Blizak, Chafiqi & Kendil, 2009; Epik et al., 2002; Galili & Hazan, 2000; Taşlıdere & Eryılmaz, 2015), and factors affecting the shadow formation (Galili & Hazan, 2000), and a full shadow. On the other hand, the most basic concepts in the light diffusion unit are the diffusion and reflaction of light. In this context, the light, which is an integral part of our daily life, and connected concepts are fully and correctly must understand. However, studies display that student's experience mental confusion about light and have low levels of conceptual understanding due to the missing experiences they have in daily life (Ayvacı & Candaş, 2018; Şahin, İpek & Ayas, 2008). In the results of Rasch analysis for LCUT, the ability level of question items coded as S3 (about normal of the surface), S4 and S6 (about reflaction of light on smooth and rough surfaces) are below the average difficulty level. Inadditon, the ability level of question item coded as S2 (about the laws of reflaction) is above the average difficulty level and is challenges for students. These results support the findings of the studies in the literature, which students have a lack of knowledge about the reflaction of light and the laws of reflaction (Aydoslu, 2018, Fariyani, Rusilowati & Sugianto, 2017; Kaplan, 2017; Wahyuningsih, Rusilowati & Hindarto, 2017). For example, Fariyani, Rusilowati, & Sugianto (2017) and Wahyuningsih, Rusilowati, &

Hindarto (2017) determined that students with a low level of conceptual understanding have alternative concepts in terms of laws of reflaction, incoming beam, reflected beam, and normal of surface.

If light encounters a substance, the substances are classified as transparent, translucent, and opaque according to their light transmission. However, the results of study point out that student has lack knowledge in the context of classification of substances to light transmittance. As a matter of fact, in the results of Rasch Analysis, the items coded as S7 and S8 (about light transmission status) are the questions on average item difficulty, but below the average person's ability. Accordingly, we conclude that the students' conceptual understanding level for item coded as S7 and S8 is at a medium level. This result of the study coincides with the studies in the literature and indicates that students have difficulty in classifying substances according to light transmission (Kaplan, 2017). On the other hand, the ability level of question items coded as S1 and S10 (about the diffisuon of light) indicates that students are at a very high level of conceptual understanding.

RM is a theory-based approach to scale development through hypothesis testing (Clements, Sarama & Liu, 2008). The construct validity of the measurement tool is ensured when the data fit RM. To answer the first question of the study, we examined infit and outfit mnsq, point measurement correlation, and reliability statistics. As a result, we observed that these statistical values were within the desired limits. Accordingly, LCUT is a useful measurement tool with measurement reliability and validity that allows measuring the level of conceptual understanding of light. In conclusion, we determined that students' level of conceptual understanding toward the concept of light is above medium level. Overall, the results on the conceptual understanding level by Rasch analysis for LCUT are coherent with studies related to conceptual understanding on light in literature and so, will contribute to the field with this aspect.

Possible Research Limitations

There are some potential limitations of the current study. First, RM was used instead of CTT in determining the level of conceptual understanding, but only fifth-grade students and fith-grade Light Unit were preferred for rasch analysis. The results indicate that the measurement tool developed for determining the conceptual understanding level is a valid and reliable measurement tool according to RM, and that the conceptual understanding level can be determined by RM. However, it would be appropriate to compare the analyses with the Light unit and other units, and at other grade levels regarding the effectiveness of RM. On the other hand, multi-tier tests in the literature mostly diagnose misconceptions and present them to the educator. However, in the study, we examined the test parameters of RM on LCUT and so did not diagnose misconceptions. Therefore, the data of the study were limited by the statistical parameters of RM, and RM was not used to determine misconceptions.

Suggestions for Future Research

In line with the findings and comments of the research, we suggested in the following:

- Should be given priority to the use of the analyses that is free from the limitations of CTT, conducted within the framework of LTT instead of the analyses based on CTT, as it gives clearer results during the evaluation of the participants.
- Increasing the use of Wright Maps in evaluations based on Rasch analysis will provide satisfying results in terms of evaluation of tests.
- Analysis results based on RM can be confirmed with exploratory and confirmatory factor analyses in the future research.
- The frequent use of multi-tier tests, such as four-tier tests, where the participants can be assessed about how confident they are with their answers rather than multiple-choice tests, will provide more accurate and reliable results in the diagnosis of conceptual understanding and misconceptions.

- It will be beneficial to employ various methods and techniques in order to make the knowledge gained by students' daily life activities permanent in subjects such as light, which is an integral part of daily life.
- Since the data obtained from multi-tier tests are used in the diagnosis of misconceptions, it will be appropriate to conduct the analyses based on RM regarding possible misconceptions.

Disclosure Statement
No potential conflict of interest was reported by the authors.

Data Availability Statement
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

REFERENCES

Akın, Ö. & Baştürk, R. (2012). The evaluation of the basic skills in violin training by many facet rasch model. *Pamukkale University Journal of Education, 31(*31*),* 175-187.

Altun, D. G. (2006). *Effect of sound and light unit prepared using multiple intelligence theory to the student success, level of retaining, attitudes towards science and multiple intelligence theory*. (Unpublished master's thesis). Institute of Science, Gazi University, Ankara, Turkey

Aminudin, A. H., Kaniawati, I., Suhendi, E., Samsudin, A., Coştu, B., & Adimayuda, R. (2019). Rasch analysis of multitier open-ended light-wave ınstrument (MOLWI): Developing and assessing second-years sundanese-scholars alternative conceptions. *Journal for the Education of Gifted Young Scientists*, *7*(3), 557-579. doi:10.17478/jegys.574524

Andersson, B., & Bach F. (2005). On designing and evaluating teaching sequences taking geometrical optics as an example. *Science Education, 89*, 196-218. doi:10.1002/sce.20044

Anshel, M. H., Weatherby, N. L., Kang, M., & Watson, T. (2009). Rasch calibration of a unidimensional perfectionism inventory for sport. *Psychology of Sport and Exercise*, *10*(1), 210-216. doi:10.1016/j.psychsport.2008.07.006

Apaydın, Z., Akman, E., Taş, E., & Peker, E. A. (2014). Analysis of knowledge structures about light concept of first level elementary students according to conceptual change theories. *Journal of Computer and Education Research*, *2*(3), 44-68. Retrieved from https://dergipark.org.tr/tr/pub/jcer/issue/18615/196501

Aydoslu, M. (2018). *Determination of cognitive structures and misconceptions of middle school students about the concepts of light and reflection by using alternative measurement and evaluation techniques*. (Unpublished master's thesis). Institute of Science, Kırıkkale University, Kırıkkale, Turkey

Ayvacı, H. S., & Candas, B. (2018). Students' understandings on light reflection from different educational level. *Journal of Computer and Education Research*, *6*(11), 1-32. doi:10.18009/jcer.309748

Baharun, N., Razi, N. F. M., Abidin, R. Z., Musa, N. A. C., & Mahmud, Z. (2017). Measuring students' understanding in counting rules and its probability via e-learning mode: a Rasch measurement approach. *Journal of Fundamental and Applied Sciences*, *9*(6S), 429-441. doi:10.4314/jfas.v9i6s.33

Benek, İ., & Kocakaya, S. (2012). Students' opinion on learning in stations technique. *Journal of Research in Education and Teaching*, *1*(3), 2146-9199. Retrieved from http://www.jret.org/FileUpload/ks281142/File/02z.benek.pdf

Blizak, D., Chafiqi, F., & Kendil, D. (2009). Students misconceptions about light in Algeria. In *Education and Training in Optics and Photonics* (p. EMA5). Optical Society of America. doi:10.1364/etop.2009.ema5

Bond, T. G & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Inc. Publishers

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE-Life Sciences Education*, *15*(4), rm4.1-7. doi:10.1187/cbe.16-04-0148

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1416898. doi:10.1080/2331186x.2017.1416898

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research, utilizing multiple choice tests. *Science Education*, *90*(2), 253-269. doi:10.1002/sce.20106

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self efficacy data. *Science Education*, *95*(2), 258-280. doi:10.1002/sce.20413

Bulut, G. (2018). Açık ve uzaktan öğrenmede şans başarısı: Klasik test kuramı (KTK) ve madde tepki kurama (MTK) temelinde karşılaştırmalı bir analiz. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, *4*(1), 78-93. Retrieved from https://dergipark.org.tr/tr/pub/auad/issue/35189/390471

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2014). *Scientific research methods*. Ankara: Pegem Publications.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The research based early maths assessment. *Educational Psychology*, *28*(4), 457-482. doi:10.1080/01443410701777272

Çetin, Ş. (2019, November). Analysis of open-ended questions with many facet Rasch model. In 4th International Symposium on Innovative Approaches in Health and Sports Sciences. Samsun, Turkey

Demirci, N., & Ahçı, M. (2016). University students' conceptual understanding on the subjects of light and optics. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education* *10*(1), 142-181. doi:10.17522/nefefmed.39726

Demirtaşlı, N. Ç. (1996). New approaches in test development: The latent trait theory-basic features, assumptions, models and limitations. *Ankara University Journal of Faculty of Educational Sciences*, *28*(2), 161-173. doi:10.1501/Egifak_0000000298

Djanette, B., Fouad, C., & Djamel, K. (2013). What thinks the university's students about propagation of light in the vacuum? *European Scientific Journal*, *9*(24), 197-213.

Doğan, N. & Tezbaşaran, A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemler bağlamında karşılaştırılması [Comparison of classical test theory and latent traits theory by Samples]. *Hacettepe University Journal of Ecucation*, *25*(25), 58-67.

Doğru, Ş. C. (2019). *The comparison of classical test theory and Rasch model in determining the psychometric properties of mixed tests.* (Unpublished master's thesis). Institute of Science, Hacettepe University, Ankara, Turkey

Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators (2003). Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950-963. doi:10.1016/s0003-9993(03)00035-2

Eggert, S., & Bögeholz, S. (2010). Students' use of decision making strategies with regard to socioscientific issues: An application of the Rasch partial credit model. *Science Education*, *94*(2), 230-258. doi:10.1002/sce.20358

Epik, Ö., Kalem, R., Kavcar, N., & Çallıca, H. (2002). Investigation of student views on the concepts of light, image formation and image observation. *Buca Faculty of Education Journal*, *14*, 64-73

Fariyani, Q., Rusilowati, A., & Sugianto, S. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, *6*(3), 1724-1729. doi:10.15294/USEJ.V6I3.20396

Fisher, W.P. Jr (2007). Rasch measurement transaction. *Rasch Measurement Transactions 21*(1), 1087-1096

Galili, I., & Hazan, A. (2000). Learners' knowledge in optics: interpretation, structure and analysis. *International Journal of Science Education*, *22*(1), 57-88. doi:10.1080/095006900290000

Gülkaya, D. (2018). *Rasch analysis and its application*. (Unpublished master's thesis). Institute of Science, Süleyman Demirel University, Isparta, Turkey

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47. doi:10.1111/j.1745-3992.1993.tb00543.x

Irawan, A. G., Nyoman Padmadewi, N., & Artini, L. P. (2018). Instructional materials development through 4D model. In *SHS Web of Conferences*, *42*. doi:10.1051/shsconf/20184200086

İlhan, M., & Güler, N. (2017). Testing the agreement between ability estimations made through classical test theory and the ones made through Rasch analysis in likert type scales. *Ege Journal of Education*, *18*(1), 244-265. doi:10.12984/egeefd.289576

Kan, A. (2006). An empirical study on the comparison of predicted item parameters with respect to classical and item response test theories. *Mersin University Journal of the Faculty of Education*, *2*(2), 227-235. Retrieved from https://dergipark.org.tr/tr/pub/mersinefd/issue/17389/181743

Kaplan, E. (2017). *Determination of sixth grade students misconceptions on, the light and sound unit with concept test, concept cartoons, and semi-structured interviews.* (Unpublished master's thesis). Graduate School of Educational Sciences, Erciyes University, Kayseri, Turkey

Kara, İ. Avcı, D., & Çekbaş, Y. (2008). Investigation of the science teacher candidates' knowledge level about the concept of light. *Mehmet Akif Ersoy University Journal of Education Faculty, 8* (16), 46-57

Kaskatı, O. (2011). *Development of computer adaptive testing method using with Rasch models for assessment of disability in rheumatoid arthritis patients*. (Unpublished doctorate thesis). Institute of Health Science, Ankara University, Ankara, Turkey

Kauertz, A., & Fischer, H. E. (2006). Assessing students' level of knowledge and analysing the reasons for learning difficulties in physics by Rasch analysis. In X.Liu &W.J.Boone (Eds.) *Applications of Rasch measurement in science education*, (pp. 212-246). Maple Grove.MN: Jam Press

Keeves, J. P. (1998). *Educational research, methodology and measurement: An international Handbook*. Oxford: Pergamon Pres

Kelecioğlu, H. (2001). Relationship between b and a parameters in latent trait theory and p and r statistics in classical test theory. *Hacettepe University Journal of Education*, *20,* 104-110

Linacre, J.M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16(2),* 878. Retrieved from https://www.rasch.org/rmt/rmt162f.htm

Linacre, J. M. (2006). Data variance explained by Rasch measures *Rasch Measurement Transactions*, *20*(1), 1045. Retrieved from https://www.rasch.org/rmt/rmt201a.htm

Linacre, J. M. (2014). Reliability and separation of measures. *A user's guide to Winsteps Ministep Rasch-model computer programs (version 3.81.0).* Retrieved from http://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (2019). *A user guide to winsteps ministep Rasch model computer programs: Program manual 4.4.7*. Retrieved from https://www.winsteps.com/a/Winsteps-Manual.pdf

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. IAP Information Age Publishing

Maat, S. M. (2015). Psychometric evaluation on mathematics beliefs instrument using Rasch model. *Creative Education*, *6*(16), 1797-1801. doi:10.4236/ce.2015.616183

Mazlum, E. & Yiğit, N. (2017). Examining indicators of knowledge of light concept through peer tutoring applications. *Hacettepe University Journal of Education*, *32*(2), 295-311. doi:10.16986/HUJE.2016019933

Mešić, V., Neumann, K., Aviani, I., Hasović, E., Boone, W. J., Erceg, N., Grubelnik, V., Sušac, A., Glamočić, D.S., Karuza, M., Vidak, A., Alihodžić, A., & Repnik, R. (2019). Measuring students' conceptual understanding of wave optics: A Rasch modeling approach. *Physical Review Physics Education Research*, *15*(1), 010115(20). doi:10.1103/physrevphyseducres.15.010115

Ministry of Education (MNE) (2018). *Primary education institutions (primary and secondary schools) science course (3,4,5,6,7,8. Grades) curriculum*. Ankara: Ministry of Education, Board of Education and Discipline

Othman, N. B., Salleh, S. M., Hussein, H., & Wahid, H. B. A. (2014). Assessing construct validity and reliability of competitiveness scale using Rasch model approach. In *Proceedings of the 2014 WEI International Academic Conference* (pp. 113-120)

Özcan, Ö., & Tavukçuoğlu, E. (2018). Investigating the high school students' cognitive structures about the light concept through word association test. *Journal of Education and Future*, (13), 121-132. doi:10.1063/1.5025996

Planinic, M., Ivanjek, L., & Susac, A. (2010). Rasch model based analysis of the force concept inventory. *Physical Review Special Topics-Physics Education Research*, *6*(1), 010103(20). doi:10.1103/physrevstper.6.010103

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 1-14. doi:10.1103/physrevphyseducres.15.020111

Preece, P. F. W. (1979). Objective measurement in education: The Rasch model. *School Science Review*, *60*(213), 770-73. ERIC ID. EJ207029

Rasch, G. (1961, July). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability.* Berkeley, USA.

Siang, L. K. (2011). Developing a profile of conceptual understanding and misconceptions in Newtonian mechanics: Rasch modelling approach. In *International Conference on Measurement and Evaluation in Education, University of Malaya*. Kuala Lumpur, Malaysia.

Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, *98*(4), 581-613. doi:10.1002/sce.21118

Şahin, Ç., İpek, H., & Ayas, A. (2008, June). Students' understanding of light concepts primary school: A cross-age study. *Asia-Pacific Forum on Science learning and teaching*, *9*(1), 1-119. Retrieved from https://www.eduhk.hk/apfslt/

Şenel, S.Ö. (2016). *Activities based on multiple intelligence theory related on learning and effects about the 7th grade subject light.* (Unpublished master's thesis). Gazi University, Ankara, Turkey

Taşlıdere, E. & Eryılmaz, A. (2015). Assessment of pre-service teachers' misconceptions in geometrical optics via a three-tier misconception test. *Bartın University Journal of Faculty of Education, 4* (1), 269-289. doi:10.14686/buefad.2015111057

Uzun, E. & Karaman İ. (2015). An analysis of prospective science teachers' mental models about light and sound. *Ekev Akademi Dergisi. 20*(65), 141-154. doi:10.17753/Ekev549

Uzunsakal, E. & Yıldız, D. (2018). A comparison of reliability tests in field researches and an applicatıon on agricultural data. *Applied Social Sciences Journal of Istanbul University-Cerrahpasa*, *2*(1), 14-28. Retrieved from https://dergipark.org.tr/en/pub/iuusbd/issue/38311/399621

Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson's classroom test of scientific reasoning. *Physical Review Physics Education Research*, *14*(2), 020104(14). doi:10.1103/physrevphyseducres.14.020104

Wahyuningsih, S., Rusilowati, A., & Hindarto, N. (2017). Analysis of misconception to science literacy using three tier multiple choice test in the materials of characteristic of light. *Unnes Science Education Journal*, *6*(3), 1736-1743. doi:10.15294/USEJ.V6I3.20426

Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch measurement to validate the instrument of students' understanding of models in science (sums). *International Journal of Science and Mathematics Education*, *12*(5), 1067-1082. doi:10.1007/s10763-013-9459-z

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, *89*(3), 335-345. doi:10.1021/ed100852t

Yüksel, S. (2012). *Analyzing differential item functioning by mixed rasch models which stated in scales*. (Unpublished doctorate thesis). Institute of Health Education, Ankara University, Ankara, Turkey

Zain, J. M., Mohd, W. M. W., & El-Qawasmeh, E. (Eds.). (2011). *Software engineering and computer systems*. In Proceedings of Second International Conference, ICSECS 2011, Kuantan, Pahang, Malaysia