



Effect of Extreme and Acquiescence Response Style in TIMSS 2015*

Munevver ILGUN DIBEK¹

ARTICLE INFO

Article History:

Received: 26 Nov. 2019

Received in revised form: 11 Feb. 2020

Accepted: 14 Feb. 2020

DOI: 10.14689/ejer.2020.87.10

Keywords

acquiescence response style, cross-cultural study, extreme response style, TIMSS

ABSTRACT

Purpose: Cross-cultural comparisons based on ordinal Likert-type rating scales have been threatened by response style which is systematic tendencies to respond to items regardless of the item content. So, this study aimed to investigate the effect of extreme response style and acquiescence response style on TIMSS 2015 data.

Method: The sample of this descriptive study

included eighth grade students of the countries Japan, Korea, Taipei, Turkey, Oman and Jordan. Students' responses to scale regarding value on mathematics were used. To examine the impact of response styles, partial credit model and partial credit model with response style were analyzed. Also, the estimates obtained from these models were compared

Findings: It was found that response styles existed in TIMSS 2015 data. Furthermore, the number of the students selecting the extreme categories were found to be lower than that of the students selecting relatively middle response categories. Additionally, item thresholds of the extreme categories were found to be distorted leading to biased determination of item response curves.

Implications for Research and Practice: The presence of the response style in the large-scale assessment which guides policy makers in their regulations in the educational systems and gives information to teachers in their practices lead researchers to examine and control the effect of them.

© 2020 Ani Publishing Ltd. All rights reserved

* This study was partly presented at the 6th International Eurasian Educational Research Congress in Ankara, 19-22 June, 2019

¹ TED University, Faculty of Education, TURKEY, e-mail: munevver.ilgun@tedu.edu.tr, ORCID: <https://orcid.org/0000-0002-7098-0118>

Introduction

Political and social scientific awareness of the globalizing world has shaped the trends of the topics addressed in the research studies. More precisely, they have boosted the cross-cultural studies to focus on non-cognitive constructs in recent decades thanks to their ability to predict not only cognitive ability but also educational and organizational outcomes (Hough & Dilchert, 2010). Also, focusing on non-cognitive variables in education and organizational research might give a chance to better predict the achievement in these areas and help to understand these variables in cultural contexts. Especially in education, there is an increased interest of cross-cultural studies in examining non-cognitive factors and their relationships with educational achievement outcomes (Richardson, Abraham & Bond, 2012). Despite many advantages, assessment of non-cognitive constructs such as value and attitude have a number of handicaps that are not the case for cognitive ability assessment. One of them is that test scores obtained from assessment of non-cognitive constructs may be susceptible to the influence of response styles (McGrath, Mitchell, Kim & Hough, 2010). The primary approach used to measure non-cognitive characteristics is to provide a set of statements with a sequential list of descriptors to respondents to determine their level of agreement (Likert, 1932). However, using the list of descriptors has been reported to be vulnerable to response style bias such as extreme response style (ERS), acquiescence response style (ARS), and mid-point response style, etc. (Van Herk, Poortinga & Verhallen, 2004). They threaten the validity of the scores obtained from the scales measuring non-cognitive constructs (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). In cross cultural research, the most commonly encountered response styles affect the associations between a construct and the substantive trait of interest are ARS and ERS (Fischer, 2004). ARS is the tendency to provide positive response to the items without considering their content (Van Herk, Poortinga & Verhallen, 2004). It is also called as "agreement tendency" (Greenleaf, 1992). On the other hand, ERS is the tendency to select extreme end points of response categories such as "strongly agree/disagree" (Chun, Campbell & Yoo, 1974). Specifically, differences in ERS could distort differences in the group means. It also affects item correlations and increases or decreases reliability regarding the internal consistency. Moreover, ERS affects the level of correlations between measures, and thus the results obtained from factor or cluster analyses. On the other hand, ARS usually causes the mean of the item to be estimated more or less than its' true score (Fischer, Fontaine, van de Vijver & van Hemert, 2009), leading to biased results. These biased scores may lead to Type I or Type II error, resulting in erroneous conclusions (Hutton, 2017). Since both attitudes and the response style can differ from one culture to another, obtaining the difference among these attitudes can either present the actual cultural differences in attitudes of the interest or in response style (Eid, Langeheine & Diener, 2003). Therefore, to reveal the real situation which exist in the different cultures, the impact of response styles should be examined.

In the literature, there is no single accepted method addressing response style threat although there is a consensus that they adversely affect the measurement of attitudes. This division in approaches may prevent applied researchers investigating

the degree to which measurement issues distort their findings and controlling for such biases systematically. In this regard, this paper comes up with several important arguments to use Item Response Theory (IRT) approach for detecting ERS and ARS in culturally diverse groups which allows for adjustment for response styles. To make social researchers more familiar with the issue of detecting response style, Tutz, Schauberger and Berger (2018) proposed a Partial Credit model with response style (PCMRS) which allows for adjustment for response style behavior. Usefulness in differentiating the substantive trait and response style and being easy to implement in respective software make this model readily available to researchers.

Alternative Approaches

In general, in the literature, two different approaches exist for handling the response styles. In the first one, items that are uncorrelated with the items measuring the substantive characteristic are (Greenleaf, 1992; Weijters, Geuens & Schillewaert, 2010) added to scale to detect the response styles. In the second one, the scale's own items which were originally intended to measure the substantive characteristics are used. In other words, in this approach there are no extra items added to the scale. A disadvantage of them is that "they are generally little to rectify the effects of response style on resulting scores once detected" (Bolt & Johnson, 2009, p.337). In other words, they do not allow researchers to obtain response style-adjusted scores of the individuals.

In addition to different approaches detecting response styles, various statistical techniques were used to examine them. The most primitive one is to determine several descriptive statistics (Reynolds & Smith, 2010). This approach is simple when compared to the other approach. However, descriptive statistics are not sufficiently explanatory enough since this technique cannot distinguish the response styles from the trait of interest. Therefore, it is difficult to determine whether the responses of the individuals reflect the response style, the actual characteristic to be measured, or both.

Other than primitive techniques, there are also more novel techniques which were introduced in the framework of Structural Equation Modeling (SEM) or IRT. In the first technique, confirmatory factor analysis (CFA) was performed to detect response styles. In CFA, response styles were usually considered as continuous latent variables (Billiet & McClendon, 2000). Instead of using CFA, latent class analysis can be used to determine subgroups of individuals who show different behaviors in terms of choosing the response categories. However, at this time the response styles were handled as categorical variables (Moors, 2010). In the second technique, several studies used a multidimensional nominal response model to determine and adjust the effect of ERS (Bolt & Johnson, 2009). Moreover, PCM, one of the polytomous IRT models, was adapted as mixture models to determine latent groups of different response styles (e.g., Austin, Deary & Egan, 2006). In mixture models, it was supposed that respondents belong to distinct latent classes. While some of the classes may represent response styles, some of them may represent the substantive trait. In this case, from one class to other class item response models fitting within different classes can change. A problem of performing them is that the number of classes is not known in

advance. Therefore, how to interpret the model within classes is a problematic issue. Even if the number of classes is determined, it is still hard to explain the difference between classes. Also, explanation of the trait represented by a class can be more complicated since it may be a response style or another trait responsible for selection of item response categories. Additionally, response styles are considered as a discrete trait (Bolt & Johnson, 2009). However, in the psychology, response style is often considered as a continuous trait (e.g. Prediger, 1999). In this case alternative models will be more proper. Recently, item response tree (IRTtree) models were analyzed to investigate response styles (Böckenholt & Meiser, 2017; Ilgun Dibek, 2019). It is more flexible in terms of modelling item response data. Also, it provides the researcher to model different types of response styles separately. However, flexibility brings along additional difficulties. Constructing the correct tree is sometimes difficult because there may be many options. On the other hand, there is no such vague situation when PCMRS, which is based on IRT framework, is used to model response style of the individuals. Also, it enables to determine whether response style exists or not. Furthermore, if the response style exists in data set, PCMRS allow to determine how strong the response style is (Tutz, Schauburger & Berger, 2018). PCMRS is distinct from all these strategies. In PCMRS, a specific parameterization is used. More specifically, for each individual, an additional parameter that determine the tendency of the individual who select extreme or middle response categories is added into the model. The general advantage of the PCMRS model for ERS and ARS lies in its simplicity of calculation and usefulness in clarifying several essential questions of these response styles, especially the amount of variance in person parameters that the response styles accounted for as well as the effect of them on estimates of item parameters (Tutz et al., 2018). As opposed to mixture models, this model provides the researchers to handle the response style as a continuous trait. Also, in this model, parameters regarding ability and response style can be simultaneously estimated, which helps to determine to the relationship between them. This approach can be used with not only partial credit model but also with ordinal latent trait models (Tutz et al., 2018). In fact, PCMRS can be seen as an extended version of PCM. To explain the association between PCM and PCMRS, after the basic PCM is briefly explained, PCMRS that includes response style parameters explicitly is considered.

The Partial Credit Model

Masters (1982) introduced the PCM. Suppose that the response of person p on one specific item i is given by $Y_{pi} \in \{0,1,\dots,k\}$. In PCM, the probability of selecting the response category " r " is as follows:

$$P(Y_{pi} = r) = \frac{\exp(\sum_{l=1}^r \theta_p - \delta_{il})}{\sum_{s=0}^k (\sum_{l=1}^s \theta_p - \delta_{il})}, \quad r = 1, \dots, k,$$

In this equation, θ_p denotes the person parameter regarding substantive trait and $(\delta_{i1}, \dots, \delta_{ik})$ denotes item parameters of item i . If one considers adjacent categories $(r - 1, r)$, PCM model can be presented as

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p - \delta_{ir}, \quad r=1, \dots, k.$$

The Partial Credit Model with Response Style

Let the categories $0, \dots, k$ denote the graded response categories of an item. The number of response categories that the item has can be even or odd:

odd number of response categories. If there is an odd number of response categories, then k is even, and assume that m represents the mid-point of the response categories (i.e. $m=k/2$). In PCM, the predictor, when selecting between categories $r-1$ and r , is denoted by $\eta_{pir} = \theta_p - \delta_{ir}$. The parameter δ_{ir} identifies the choice between categories $r-1$ and r . ARS and ERS are modeled by adjusting the thresholds δ_{ir} . To detect the effect of response style, one more person parameter γ_p is added in the predictor. This parameter moves the thresholds of categories representing agreement and disagreement in opposite directions. In this case, PCMRS can be formulated as (Tutz, Schauberger & Berger, 2018):

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p + \gamma_p - \delta_{ir}, r = 1, \dots, m$$

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p - \gamma_p - \delta_{ir}, r = m+1, \dots, k$$

In PCMRS, the predictor, when selecting between categories $r-1$ and r , is as follows:

$$\eta_{pir} = \theta_p + \text{sgn}(m - r + 0.5)\gamma_p - \delta_{ir}, r = 1, \dots, k$$

where $\text{sgn}(\cdot)$ represents the sign function. When " x " takes value greater than "0" the $\text{sgn}(x)$ takes a value of "1", and when x takes value greater than "0", the $\text{sgn}(x)$ takes a value of "-1". Lastly, if $x=0$, then $\text{sgn}(x) = 0$. In this case, the response categories are divided into three categories. These categories are categories indicating the disagreement, the neutral category and categories indicating the agreement of the participants.

even number of response categories. If the number of categories is even, then k is odd. In this case, the response categories are divided into the disagreement and agreement categories at the point $m = [k/2] + 1$. So, the related PCMRS addressing the tendency of choosing middle and extreme categories can be formulated as follows:

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p + \gamma_p - \delta_{ir}, r = 1, \dots, m-1$$

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p - \delta_{ir}, r = m$$

$$\log \left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p - \gamma_p - \delta_{ir}, r = m+1, \dots, k$$

For this case, the predictor can be defined as $\eta_{pir} = \theta_p + \text{sgn}(m-r)\gamma_p - \delta_{ir}$, $r=1, \dots, k$.

To sum up, it is clearly understood that PCMRS model allows researchers to determine the effect of ERS and ARS simultaneously and it can be used for both even numbered and odd numbered response categories, as well. This paper is built on the study of Tutz, et.al.,(2018). The present study makde contributions to the related literature in many ways. Firstly, it provided the reader with a general picture of alternative approaches for detecting ARS and ERS in survey data. Moreover, in this study, in addition to brief explanation of partial credit model (PCM), a detailed explanation of the PCMRS proposed by Tutz, Schauburger and Berger (2018) for determining and eliminating the effect of response style behavior in various cultures was given. Also, this study heeded the call of several authors such as Van de Vijver and Leung (2000) and Moors (2004), and empirically examined the role of response style which distorts the measurement of attitudes by focusing on the changes in item and person parameters. As a result, this study will contribute to see the actual situation of students from different countries, which helps the related policy makers of these countries to be aware of this problem and make suitable changes in their education system. In this context, the research questions that this study sought to answer were as follows:

- i. Does the effect of response styles exist in TIMSS 2015 data?
- ii. How do the response styles affect the variability in person parameters of the countries participated in TIMSS 2015?
- iii. What is the percentage of students with different response styles in the countries participating in TIMSS 2015?
- iv. How do the response styles affect thresholds of the attitudinal items?
- v. How do item response curves differ with different amount of response style parameters?

Method

Research Design

This study is a descriptive research study regarding the detection of the effect of ERS and ARS among students and items (Johnson & Christensen, 2008). In descriptive research studies, there is no manipulation. They are conducted to provide the accurate characteristics of the individuals or the phenomenon.

Research Sample

The eighth-grade students in the countries participated in TIMSS 2015 constituted the sample of this study. Students were selected by performing two-stage stratified sampling method. In the first stage, schools were randomly chosen in accordance with their proportion in the population. In the second stage, from each of these schools at

least one class was randomly chosen. All students in these classes were included in the study (LaRoche, Joncas & Foy, 2016). Population and sample of these countries are given in Table 1.

Table 1

Population and Sample

Country	Population		Sample	
	School	Student	School	Student
Japan	10406	1162528	147	4745
Korea	3007	587190	150	5309
Taipei	931	285714	190	5711
Turkey	15583	1298955	218	6079
Oman	669	55181	300	9105
Jordan	3108	145847	254	7861

As it is clear from Table 1, while number of schools in the sample changes from 147 to 300, the number of students in the sample changes from 4745 to 9105.

To determine which countries will be selected, students' scores on one of the affective constructs were included in this study considering the effect of response style on non-cognitive constructs. So, due to the association between value on mathematics and attitude toward mathematics, countries are ranked according to the percentage of students whose value on mathematics is high. In line with this criteria, three countries with the fairly highest attitude scores and three countries with the fairly lowest attitude scores were selected to determine whether the responses of the students reflect the response style or the actual score on value in mathematics. In total, six countries were selected. The percentages of the students who had a strong positive attitude toward mathematics are given in Table 2 (Mullis, Martin, Foy & Hooper, 2016).

Table 2

Percentage of The Students Who Value Mathematics High

Country	Students highly valued mathematics (%)
Japan	11
Korea	13
Taipei	10
International Average	42
Turkey	47
Oman	59
Jordan	65

As it is clear in Table 2, while the percentages of the students highly valued mathematics in Japan, Korea and Taipei are lower than that of the international average, the percentages of the students in Turkey, Oman and Jordan are higher than that of the international average. Also, it can be further stated that majority of the students in Jordan valued mathematics high.

Research Instruments and Procedures

The TIMSS 2015 student questionnaire dataset was used to conduct analysis. Related data set for each country was obtained from the official website (<https://timssandpirls.bc.edu/timss2015/international-database/>). This questionnaire includes items measuring students' demographic information, their home environment, school climate, and several affective constructs which are supposed to be related to mathematics achievement and science achievement. Specifically, in the present study, students' responses to items measuring valuing on mathematics were considered to examine the effect of ERS and ARS on them. Students valuing mathematics is related to their external motivation and it indicates the attitude towards the significance and benefits of mathematics (Wigfield & Eccles, 2000). All in all, students' levels of agreement with nine statements for this variable were measured. These statements have four response categories ranging from "strongly agree" to "strongly disagree".

The student questionnaire takes 15–30 minutes to complete. For the selected countries, the Cronbach alpha reliability coefficients of the scores obtained from students' valuing on mathematics scale ranged from .85 to .90. These scores are higher than 0.70, indicating that the reliability values are sufficient (Nunnally, 1978). Therefore, after the sufficient reliability coefficients were determined, further analyses were performed.

Since all samples of the countries were used (i.e there is no selection from sample) and imputation techniques may affect response categories (Mooi, Sarstedt, & Mooi-Rec, 2018) selected by students, the missing values in each data set were deleted instead of assigning a value. Also, in the same manner, since outliers may be the students displaying extreme response style, they were not removed from the sample, which is crucial and the main focus for this study. For the categories of attitudinal items, a reverse coding was done so that the higher values obtained from the scales would represent positive attitude toward mathematics.

Data Analysis

To determine whether the effect of response style exists in data set of the countries, a simple PCM and PCMRS that uses modified thresholds were fitted. In both models, marginal estimation was performed since the alternative estimation methods have several handicaps. For example, the joint likelihood estimation has to estimate many parameters, which causes estimates unstable. Also, it leads to asymptotically biased estimates (Tutz, Schauburger & Berger, 2018). Before conducting the analysis, assumptions of unidimensionality, local item independence, monotonicity, invariance of item and person parameters were tested (Hambleton & Swaminathan, 1985). More

precisely, when the scree plots for each country were examined, it was seen that there was a main factor, providing evidence of unidimensionality. Additionally, the local independence assumption was also met since the unidimensionality assumption was met as stated by Hambleton and Swaminathan (1985). Also, it was determined that the probability of selecting higher response categories of the item increases as the level of the individual's ability increases, that is, the option characteristic curves increase monotonically. To test invariance of the item parameters, item parameters were estimated by using two groups of students who were selected from the sample for each country and found similar to each other. To test the fact that person parameters are free from the items, the person parameters were estimated by using two different item sets and found to be similar to each other. All in all, all assumptions were met. Person parameters for the PCM and person and response style parameters for the PCMRS were assumed normally distributed. The estimated variance of the person parameters ($\hat{\sigma}^2$) and the estimated covariance matrix

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{\theta}^2 & \hat{cov}_{\gamma,\theta} \\ \hat{cov}_{\gamma,\theta} & \hat{\sigma}_{\gamma}^2 \end{pmatrix}$$

between person and response style parameters were calculated by fitting the PCM and PCMRS to determine the presence of response styles in TIMSS 2015 data and the role of them in the variability of person parameters of the countries participated in TIMSS 2015. Additionally, by analyzing PCMRS model, for each student, the trait parameter (γ_p) regarding response style were computed to determine the percentages of the students exhibiting ERS and ARS.

To determine the effect of response styles on item parameters and item response curves, scaled shifting of thresholds were used. Since the items used in this study has four response categories, individuals have to select either agreement or disagreement categories. In this case, for example, for item 1, the parameters δ_{11} , δ_{12} and δ_{13} determining the choice between categories 1 and 2, 2 and 3 and 3 and 4, respectively are modified as $\hat{\delta}_{11} = \delta_{11} - \gamma_p$, $\hat{\delta}_{12} = \delta_{12}$, and $\hat{\delta}_{13} = \delta_{13} + \gamma_p$, where the parameters δ_{ir} are estimated by PCM. The same modifications in item parameters were done for the other items. Item and person parameters were estimated using R package of "PCMRS" (Schauberger, 2018) and item response curves were plotted by using R packages of "dplyr" (Wickham, Francois, Henry & Muller, 2019), "mirt" (Chalmers, 2012) and "mirtCAT" (Chalmers, 2016). These parameters and curves obtained by using PCM and PCMRS were compared.

Results

To determine whether the response style exist in TIMSS 2015 data and the influence of the response style on the variability in the person parameters of the countries, the estimated variance of the person parameters ($\hat{\sigma}^2$) and the standard deviation of the response style parameters for the countries were determined by fitting PCM and PCMRS and obtaining ($\hat{\Sigma}$). They were presented in Table 3:

Table 3*Estimates Obtained from PCM and PCMSRS*

Countries	$\hat{\sigma}^2$	$\hat{\Sigma}$
Japan	1.73	$\begin{pmatrix} 1.47 & .13 \\ .13 & 3.24 \end{pmatrix}$
Korea	2.48	$\begin{pmatrix} 2.24 & -.09 \\ -.09 & 5.81 \end{pmatrix}$
Taipei	2.65	$\begin{pmatrix} 2.15 & .30 \\ .30 & 4.52 \end{pmatrix}$
Turkey	1.75	$\begin{pmatrix} 1.44 & .07 \\ .07 & 2.12 \end{pmatrix}$
Oman	1.65	$\begin{pmatrix} 1.17 & .06 \\ .06 & 1.80 \end{pmatrix}$
Jordan	2.49	$\begin{pmatrix} 1.64 & -.05 \\ -.05 & 2.67 \end{pmatrix}$

As it is shown in Table 3., the magnitude of the standard deviations of the response style for the countries Japan, Korea, Taipei, Turkey, Oman and Jordan ($\hat{\sigma}_{\gamma(japan)} = 1.80$, $\hat{\sigma}_{\gamma(korea)} = 2.41$, $\hat{\sigma}_{\gamma(taipei)} = 2.13$, $\hat{\sigma}_{\gamma(turkey)} = 1.46$, $\hat{\sigma}_{\gamma(oman)} = 1.34$, $\hat{\sigma}_{\gamma(jordan)} = 1.63$, respectively) indicated that the presence of response styles in the data regarding students' value in mathematics should not be ignored for all countries. After the existence of the response styles in TIMSS 2015 was proved, the percentage of the students having extreme response style or acquiescence response style were examined. They were displayed in Table 4:

Table 4*The Percentage of The Students Displaying Response Styles*

Countries	Students displaying ERS (%)	Students displaying ARS (%)	Students displaying none of ARS and ERS (%)
Japan	38.6	42.2	19.2
Korea	35.8	40.4	23.8
Taipei	32.7	40.9	23.3
Turkey	30.6	35.5	33.9
Oman	29.2	32.6	38.2
Jordan	34.9	36.8	28.3

As it can be understood from Table 4, the percentages of the students having ARS were fairly higher than that of the students having ERS for all countries. In other words, students in each country participated in TIMSS 2015 were less likely to choose extreme categories compared to other categories. In addition, the percentage of the students who do not have none of ARS and ERS ranged from 19.2 to 38.2.

When the effect of response styles on item parameters were examined, it was found that they distorted the estimates of item thresholds. In other words, the presence of response style led to biased estimation of item thresholds. Specifically, the estimates of the item parameters for item 1 were shown in Figure 1, separately for each country.

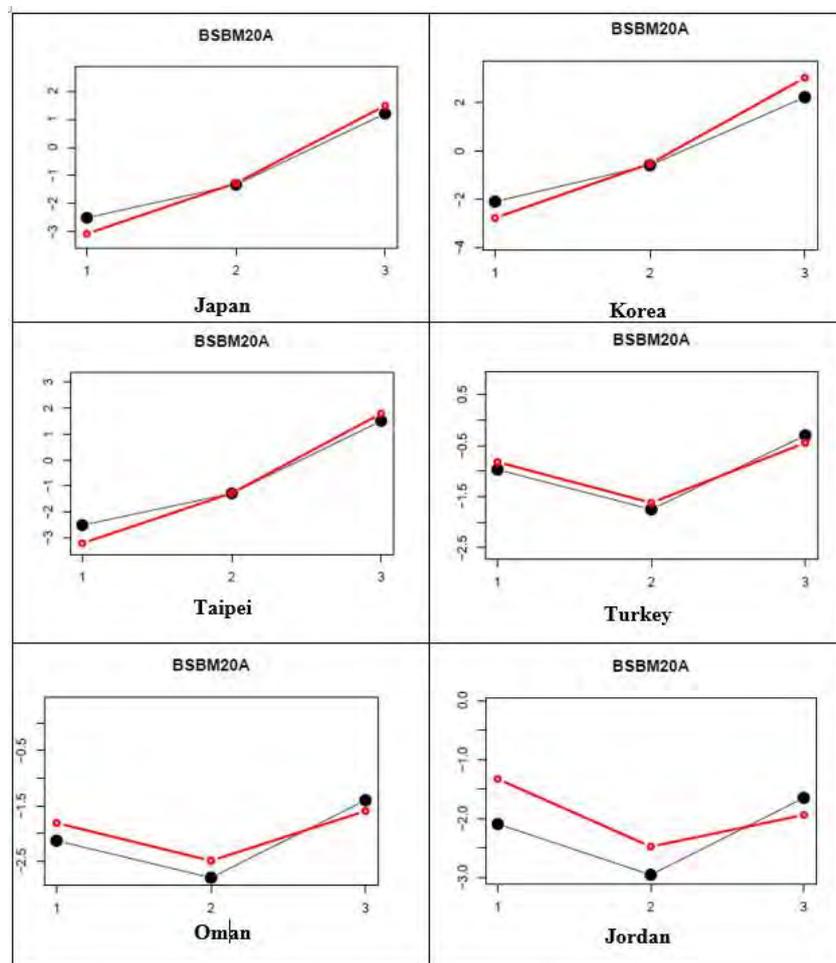


Figure 1. Estimates of thresholds for item 1 (code BSBM20A)

Since the items used in this study had four response categories, three thresholds were estimated. The red lines showed the estimates obtained from PCMRS and the black ones represented the estimates obtained from the PCM. As it is shown in Figure 1, for the first and the last threshold of the first item were obviously estimated differently by performing PCM and PCMRS for all countries, whereas the middle thresholds were fairly close to each other. In other words, it was found that when the effect of the response styles was neglected, the parameters of end points of response categories was observed to be distorted.

When the effect of different amount of response style traits (γ_p) on item response curves were analyzed, it was found that the probabilities of selecting different item categories changed depending on the value of response style parameters of the students. As an example, the item response curves obtained from the responses of the Japanese and Turkish students to students' value in mathematics scale were given in Figure 2.

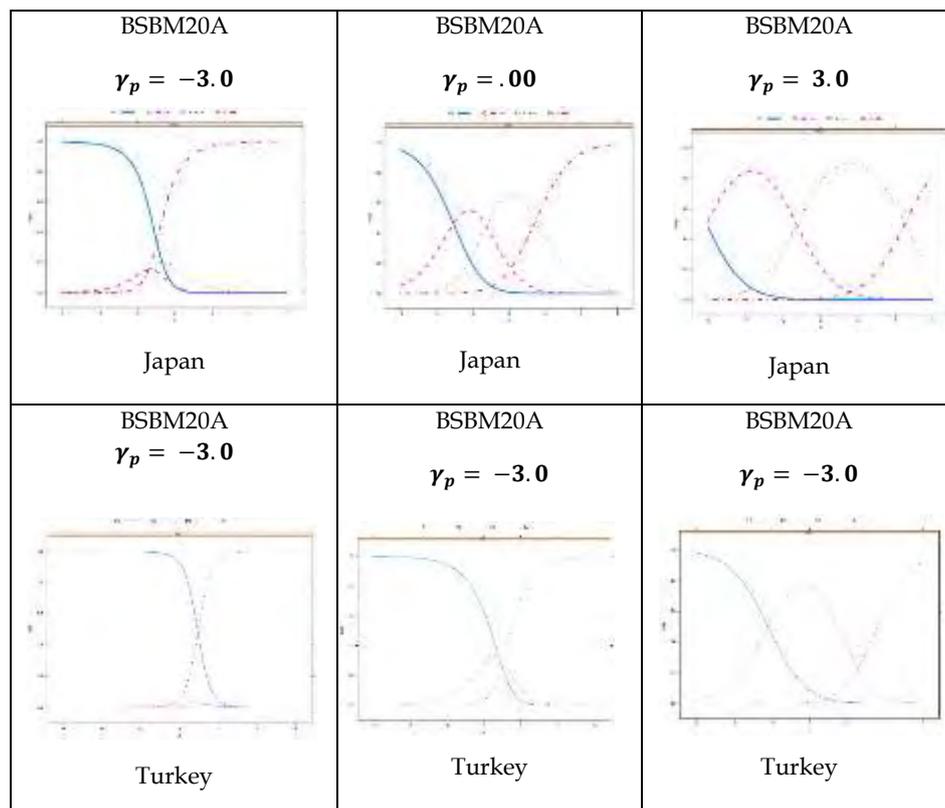


Figure 2. Item Response Curves of Item 1 (BSBM20A) with Different Gamma Parameters

It is clearly seen from Figure 2 that for both countries for $\gamma_p = -3$, the extreme categories of the item 1 (BSBM20A) were found to have higher probabilities than for $\gamma_p = 0$. The inverse was found for $\gamma_p = 3$. For example, if a Japanese student's trait regarding response style has negative value, it means that this student tended to choose more extreme categories compared to middle categories. Conversely, it was found that when the response style parameter of the Japanese student took positive value, this student was more likely to be affected by response styles ($\gamma_p = 0$), it was found that the probabilities of choosing each category were fairly close to each other. The same pattern was also observed for the other countries and other items.

Discussion, Conclusion and Recommendations

Response styles are one of the validity threats for assessment of non-cognitive constructs since they lead to biased interpretation of the differences found in international studies. Therefore, it is crucial to investigate the impact of response styles with an effective method. In this context, the current study examined the effect of ERS and ARS on students' valuing in mathematics by extending the use of a PCM model in the examination of response styles. To provide empirical evidence, both the effect of ERS and the ARS were investigated based on the responses to students' value in mathematics scale in TIMSS 2015 by including an additive parameter representing response style in PCM. To put it in different words, this study used PCMRS and calculated the estimated covariance matrix between person and response style parameters by fitting the PCM and PCMRS to determine the presence of response styles in TIMSS 2015 data.

The findings of this study replicate prior findings that response styles exist in data (Lu & Bolt, 2015). Also, it was concluded that response styles were one of the reasons of the variability in the person parameters of the selected countries participated in TIMSS 2015. The present study showed that when the effect of the response styles was not taken into consideration, variability of the person parameter regarding value on mathematics increased for each country. This finding is consistent with the study conducted by Tutz, Schauburger and Berger (2018) who investigated the effect of response styles on individuals' responses to items regarding tenseness. They found that the estimated variance of the person parameters decreased when they took into consideration the effect of response style. In this case, it can be stated that the reason for the decrease in variance within the individuals is the elimination of the difference in response style.

When the percentage of the students having ERS or ARS were examined, it was concluded that in each country the percentage of the students with ARS was higher than that of the students with ERS. This finding may result from several characteristics of the countries such as power distance, collectivism/individualism, and uncertainty avoidance (Harzing, 2006). When the cultural structures of the countries included in the current study are taken into consideration, it can be stated that the countries have a collectivistic structure according to the classification made by Hofstede (2001). Collectivistic countries prefer harmony, avoid confrontations and accept the opinions of the groups (Hofstede, 2001). They have a tendency of avoiding strong opinions.

Especially in East Asian countries, with the effect of teaching based on Confucianism, students keep themselves away from extreme decisions (Si & Cullen 1998). Therefore, societies dominated by collectivism tend to show middle or positive responses.

Concerning parameter estimation, it was concluded that response styles affected item parameters. Specifically, deviations were observed in threshold parameters at the endpoints. This effect may be associated with the variance of the latent trait regarding Japanese students' valuing on mathematics ($\theta_{p(japan)}$). The variance of this person parameter for Japan decreased from 1.73 to 1.47 when the model took into consideration the response styles. Therefore, the variability in the population is related to the response style. This finding was supported by the study of Pelieninger and Heck (2018) who investigated the effect of several response styles indicated that response styles led to biased estimation of item parameters. They further emphasized that ARS causes the item difficulty parameters of the regular items to be underestimated and that of the reverse-coded items to be overestimated. The reason for this finding may be that due to the nature of the response styles, some students tended to select some of the response categories more which yielded to the accumulation of responses at certain response categories, regardless of the scale or items' content (Pearse, 2011). Therefore, this situation results in biased estimation of item parameters.

In parallel with the previous finding related to the effect of response style on item parameters, when the effect of response styles on item characteristics curves were examined, it was concluded that depending on the magnitude of the trait regarding response styles, item category selection of the students and thus the corresponding item characteristics curve changed. In line with this finding, Bolt and Johnson (2009) indicated that individuals having high level of ERS are more likely to choose the end points of response categories as opposed to individuals having low level of ERS. They further added that item characteristic curve invariance across groups, which is one of the assumptions of traditional unidimensional IRT models, cannot be established. This finding is related to the change in the item parameters in the presence of response style, which was proved in the previous finding.

This study provides important implications for researchers or practitioners who are willing to solve validity problems in large scale surveys. This study suggests that the investigation of the possible existence of response styles should be routine when comparing different countries in terms of the affective variables that they have. The evidences presented here is sufficient to alert researchers to the possible negative effects caused by the presence of ERS and ARS. Furthermore, the finding of the current study is informative for practitioners to determine the tendency of cultures when responding the surveys. As the PCMRS model taking into consideration of response style contamination produced less variability in person parameters regarding value in mathematics, it is reasonable to indicate that differences found in cross cultural comparisons may be due to response styles. In the similar manner, policymakers should take the role of response styles into consideration while making arrangements based on international comparison results. So, it is highly recommended that they should focus not only on the effectiveness of the education system but also on such response style effects.

This study is limited in several aspects. First, the items used in this study have only four response categories available to measure ERS and ARS. Therefore, this study could not detect the effect of MRS which requires mid-point. As it has been reported that different response formats affect the existence of response styles and lead to different response styles (Hui & Triandis, 1985), it is recommended that the effect of the same response styles can be re-examined by using different item formats. Secondly, this study examined the effects of response styles on only one affective construct, and further research can be conducted with several affective constructs such as confidence, interest, etc. and personality characteristics. To sum up, this study gives valuable information about the impact of response style factor in students' self-report in TIMSS.

References

- Austin, E. J., Deary, I. J. & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235-1245. doi:10.1016/j.paid.2005.10.018
- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*, 608-628. doi:10.1207/S15328007SEM0704_5
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement, 33*, 335-352.
- Bockenholt, U. & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology, 70*, 159-181. doi:10.1111/bmsp.12086
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. doi:10.18637/jss.v048.i06
- Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory Applications. *Journal of Statistical Software, 71*(5), 1-39. doi:10.18637/jss.v071.i05
- Cheung, M. W.-L., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.
- Chun, K.-T., Campbell, J. B., & Yoo, J. H. (1974). Extreme response style in cross-cultural research: A Reminder. *Journal of Cross-Cultural Psychology, 5*(4), 465-480. doi: 10.1177/002202217400500407
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis. A primer. *Journal of Cross-Cultural Psychology, 34*(2), 195- 210.

- Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of CrossCultural Psychology*, 35(3), 263-282.
- Fischer, R., Fontaine, J. R. J., van de Vijver, F. J. R., & van Hemert, D. A. (2009). An examination of acquiescent response styles in cross-cultural research. In G. Aikaterini & K. Mylonas (Eds.), *Quod Erat Demonstrandum: From Herodotus' ethnographic journeys to cross-cultural research*. Proceedings from the 18th International Congress of the International Association for Cross-Cultural Psychology. https://scholarworks.gvsu.edu/iaccp_papers/52/
- Greenleaf, E.A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and application*. Boston: Kluwer Academic Publishers Group.
- Harzing, A.W (2006). Response styles in cross-national survey research: a 26-country study. *International Journal of Crosscultural Management*, 6(2), 1-37.
- Hofstede, G. H. (2001). *Cultures consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, California: Sage Publications, Inc.
- Hough, L., & Dilchert, S. (2010). Personality: Its measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 299- 319). New York, NY, US: Routledge/Taylor & Francis Group.
- Hutton, A.C. (2017). *Assessing acquiescence in surveys using positively and negatively worded questions* (Unpublished doctoral dissertation), Virginia Commonwealth University, Richmond, Virginia
- Hui, C. H., & Triandis, H.C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49(2), 253-60.
- Ilgun Dibek, M. (2019). Examination of the extreme response style of students using IRTree: The case of TIMSS 2015. *International Journal of Assessment Tools in Education*, 6(2), 300-313.
- Johnson, R. B., & Christensen, L. B. (2008). *Educational research: Quantitative, qualitative, and mixed approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Krosnick, J. A. (1999). Survey research. *Annual Review Psychology*, 50, 537-567.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sample Design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 3.1-3.37). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timss.bc.edu/publications/timss/2015-methods/chapter-3.html>

- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Personality*, 140, 5- 53.
- Lu, Y., & Bolt, D. M. (2015). Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style. *Large-scale Assessments in Education*, 3(2), 1-18. doi: 10.1186/s40536-015- 0012-0
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47 (2), 149-174.
- McGrath, R., Mitchell, M., Kim, B.H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136(3), 450-470.
- Mooi, E., Sarstedt, M., & Mooi-Reci, I. (2018). *Market research: The process, data, and methods using Stata*. Singapore: Springer
- Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multi-group latent class structure model with adjustment for response style behavior. *European Sociological Review*, 20(4), 303-320.
- Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, 22, 93-119. doi:10.1093/ijpor/edp036
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/international-results/>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: The case of a 21-Point scale. *The Electronic Journal of Business Research Methods*, 9(2), 159-171.
- Plieninger, H. & Heck, D.W. (2018) A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research*, 53(5), 633-654, doi: 10.1080/00273171.2018.1469966
- Prediger, D. J. (1999). Basic structure of work-relevant abilities. *Journal of Counseling Psychology*, 46, 173-184.
- Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, 13, 230-243.

- Richardson, M.D., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138 2, 353-87
- Schauberger, G. (2018). *PCMRS: Model response styles in partial credit models*. R package version 0.1-1. <https://CRAN.R-project.org/package=PCMRS>
- Si, S. X., & Cullen, J. B. (1998). Response categories and potential cultural bias: Effects of an explicit middle point in cross-cultural surveys. *International Journal of Organizational Analysis*, 6, 218-230
- Tutz, G., G. Schauberger, and M. Berger (2018). Response styles in the partial credit model. *Applied Psychological Measurement*, 42, 407-427.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data-analysis for cross-cultural research* (Vol. 1). Thousand Oaks, California: Sage Publications.
- Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31(1), 33-51.
- Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from 6 EU countries. *Journal of Cross-Cultural Psychology*, 35(3), 346-360.
- Wickham, H. Francois, R., Henry, L. and Muller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.1. <https://CRAN.R-project.org/package=dplyr>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, 34(2), 105-121. doi: 10.1177/0146621609338593

Uç ve Kabullenici Tepki Stilinin TIMSS 2015'teki Etkisi

Atf:

Ilgun Dibek, M. (2020). Effect of extreme and acquiescence response style in TIMSS 2015. *Eurasian Journal of Educational Research*, 87, 199-220. DOI: 10.14689/ejer.2020.87.10

Özet

Problem Durumu: Globalleşen dünyada politik gelişmelerle birlikte bilimsel olmayan yapıların belirlenmesine odaklanılmıştır. Bu durumun nedenleri olarak bilişsel yapılar üzerinde etkisinin olması, başarının kestirilmesinde önemli bir rolünün olması, bilişsel yapıların çeşitli bağlamlar ve kültürlerde anlaşılmasını sağlaması sıralanabilir. Özellikle kültürler arası karşılaştırma çalışmalarında, akademik öz-yeterlilik, duyuşsal zeka, tutum gibi çeşitli bilişsel olmayan yapıların ve bu yapıların başarı ile ilgili çıktılarla ilişkisi üzerindeki ilgi giderek artmaktadır. (Richardson, Abraham & Bond, 2012). Bilişsel olmayan yapıların ölçülmesinin avantajlarının yanı sıra, değer, tutum gibi yapıların ölçülmesinde bilişsel yapıların ölçülmesinde söz konusu olmayan bazı sınırlılıklar söz konusudur. Bunlardan biri bu yapıların tepki stillerinin etkisine maruz kalmasıdır (McGrath, Mitchell, Kim, & Hough, 2010). Bilişsel olmayan yapıların ölçülmesinde sıklıkla kullanılan yaklaşım, cevaplayıcılara katılım düzeylerini belirleyecekleri birtakım ifadeler listesi vermektedir. Fakat bu yaklaşım, uç tepki stili (UTS), kabullenici tepki stili (KTS), orta nokta tepki (OTS) stili gibi bazı tepki stillerinin etkisine açıktır (Van Herk, Poortinga, & Verhallen, 2004). Kültürler arası karşılaştırma çalışmalarında sıklıkla karşılaşılan tepki stilleri UTS ve KTS'dir. UTS grup ortalamaların farklılaşmasına, iç tutarlılık anlamında güvenilirliğin düşmesine neden olurken KTS tip II hatanın oluşmasına yol açmaktadır.

Alan yazında, bu geçerlilik tehdidinin belirlenmesine yönelik kabul edilmiş tek bir yöntem yoktur. Bu yöntemlerden bazılarında çeşitli betimsel istatistikler hesaplanmakta veya bilişsel olmayan yapının ölçülmesinde kullanılan ölçekteki maddelerle ilişkisiz ilave maddeler eklenmektedir. Fakat bu yöntemler, tepki stilinin miktarını belirlemede yetersiz kalmaktadır. Bu yöntemlerin yanı sıra gizil sınıf analiziyle de tepki stillerinin etkisi belirlenebilmektedir. Fakat, bu yöntemin en büyük sınırlılığı tepki stilini süresiz bir değişken olarak ele almasıdır. Madde tepki kuramına dayalı bazı yöntemlerde ise bu sınırlılık ortadan kaldırılmıştır. Örneğin madde tepki ağacı modellerinde çeşitli tepki stillerinin etkisi rahatlıkla modellenilebilmektedir. Fakat, analiz öncesinde oluşturulması gereken ağaç farklı bir şekilde oluşturuldu ise analiz sonuçları yanlış çıkabilmektedir. Böyle bir belirsizlik tepki stilinin etkisinin dahil edildiği kısmi puan modelinde söz konusu değildir. Bu model sayesinde tepki stilinin etkisinin miktarı belirlenebilmektedir. Tepki stilini sürekli bir değişken olarak ele alan bu modelde, bireye ve tepki stiline ilişkin parametreler eş zamanlı olarak kestirilebilmekte ve böylelikle tepki stili ve bireyin tutumu arasındaki ilişkiler belirlenebilmektedir.

Araştırmanın Amacı: Bu çalışmanın amacı, tepki stillerinin etkisinin TIMSS 2015’de uygulan matematiğe yönelik verilen değerle ilgili veri setinde etkisinin olup olmadığını ve bu etkinin öğrencilerin değer puanları ve madde parametrelerinde nasıl bir değişime yol açtığını belirlemektir.

Araştırmanın Yöntemi: Betimsel modelde bu olan bu araştırmanın örneklemini TIMSS 2015 uygulamasına katılan ülkelerden Japonya ($n_1=4745$), Kore ($n_2=5309$), Tayvan ($n_3=5711$), Türkiye ($n_4=6079$), Umman ($n_5=9105$) ve Ürdün ($n_6=7861$)’deki sekizinci sınıf öğrencileri oluşturmaktadır. Ülkelerin seçiminde matematiğe yönelik çok fazla değer veren öğrencilerin yüzdesinin en fazla ve en düşük olması durumu dikkate alınmıştır. Bir diğer ifade ile matematiğe fazla değer veren öğrencilerin fazla olduğu ve buna karşın başarıların düşük olduğu öğrencilerin yer aldığı ülkeler ile matematiğe değer veren öğrencilerin çok az olduğu ve buna karşın başarıların yüksek olduğu öğrencilerin yer aldığı ülkeler seçilmiştir.

Veri toplama aracı olarak öğrenci anketinin kullanıldığı bu çalışmada öğrencilerin matematiğe değer verme alt ölçeğine ait maddelere verilen cevaplar analiz sürecine dâhil edilmiştir. Bu doğrultuda, UTS’nin ve KTS’nin etkisini belirlemek amacıyla tepki stillerinin etkisinin dâhil edildiği kısmi puan modeli ve tepki stillerinin etkisinin dahil edilmedi kısmi puan modeli analiz edilmiştir. Tepki stillerinin birey parametreleri üzerindeki etkisini belirlemek amacıyla kovaryans matrisi ve birey parametrelerine ilişkin varyans değerleri hesaplanmıştır. Bunun yanı sıra, tepki stillerinin madde parametreleri üzerindeki etkisini belirlemek amacıyla madde eşik parametrelerinde düzeltme yapılmıştır. Madde ve birey parametrelerinin kestiriminde R programında “PCMRS” paketi (Schauberger, 2018) ve madde tepki eğrilerinin oluşturulmasında “dplyr” (Wickham, François, Henry, & Müller, 2019), “mirt” (Chalmers, 2012) and “mirtCAT” (Chalmers, 2016) paketleri kullanılmıştır.

Araştırmanın Bulguları: Araştırmanın sonuçlarına göre tepki stiline ait standart sapma değerleri ($\hat{\sigma}_{Y(japonya)} = 1.80$, $\hat{\sigma}_{Y(kore)} = 2.41$, $\hat{\sigma}_{Y(tayvan)} = 2.13$, $\hat{\sigma}_{Y(türkiye)} = 1.46$, $\hat{\sigma}_{Y(umman)} = 1.34$, $\hat{\sigma}_{Y(ürdün)} = 1.63$) öğrencilerin matematiğe yönelik değerlerine ait cevaplarında tepki stillerinin etkisinin olduğunu göstermektedir. Bunun yanı sıra, seçilen ülkelerde KTS’ye sahip öğrencilerin yüzdeki UTS’ye sahip öğrencilerin yüzdesinden fazladır. Araştırmanın bir diğer bulgusu ise tepki stillerinin madde eşik parametrelerinin ve buna bağlı olarak madde tepki eğrileri üzerinde etkisinin olduğudur. Öğrencilerin tepki stiline ait parametre değerlerinin miktarına göre tepki kategorilerini seçme olasılıklarının değiştiği bulunmuştur.

Araştırmanın Sonuçları ve Önerileri: Araştırmada TIMSS 2015’in matematiğe yönelik değerle ilgili veri setinde tepki stillerinin etkisinin olduğu, seçilen ülkelerdeki öğrencilerin matematiğe yönelik değerle ilgili puanlarındaki değişimin bir nedeninin öğrencilerin sergilemiş olduğu tepki stillerinin olduğu sonucuna ulaşılmıştır. Aynı zamanda, seçilen ülkelerdeki öğrencilerin KTS sergileme eğilimlerinin daha fazla olmasında ülkelerin kültürel yapılarının etkili olduğu ifade edilebilir. Bunun yanı sıra, tepki stillerinin özellikle uç noktalardaki eşik parametrelerinin kestiriminde ve bunlara bağlı olan madde tepki eğrilerinin oluşturulmasında yanlışlığa neden olduğu sonucuna ulaşılmıştır. İlerleyen

çalıřmalarda farklı madde formatlarının veya farklı sayıdaki tepki kategorilerin tepki stillerinin varlıęı konusundaki etkisi araştırılabilir. Bunun yanı sıra, tepki stilinin biliřsel olmayan farklı yapıların ölçülmesindeki etkisi de incelenebilir.

Anahtar Sözcükler: Kabullenici tepki stili, kültürler arası çalışma, uç tepki stili, TIMSS