



The Association Between Person-Fit and Extreme Response Style

Akif AVCU*

• Received: 06.12.2020 • Accepted: 24.11.2021 • Online First: 24.11.2021

Abstract

Extreme response style (ERS) refers to a disproportionate preference for endpoint options in Likert type or rating scale items, regardless of their content. The ERS has the potential to deteriorate the fit of the measurement model in use. In the current study, it was proposed that the use of the ERS is one possible cause of aberrant response patterns. In the context of item response theory (IRT), aberrant response patterns are also known as misfit responses, and these patterns are investigated using person-fit statistics. Accordingly, this study aimed to investigate whether the ERS is associated with misfit response patterns. For this purpose, data were collected from 727 university students with the Basic Personality Traits Inventory (BPTI) using a convenient sampling method. The findings revealed that people who use the ERS were more likely to produce misfit response patterns. The results were consistent for both the lower-tail and the upper-tail ERS use. Based on the findings, it was concluded that the ERS is one plausible reason for a person's misfit in the context of personality m. More in-depth analyses of the link between the ERS and person-fit are encouraged.

Keywords: extreme response style, person-fit, item response theory

Cited:

Avcu, A. (2022). The association between person-fit and extreme response style. *Pamukkale University Journal of Education*, 55, 27-45. doi:10.9779.pauefd. 836567

* Research Assistant, Dr., Marmara University, Atatürk Faculty of Education, ORCID: 0000-0003-1977-7592, avcuakif@gmail.com

Introduction

Self-report measurement instruments have been widely used as data collection methods in a variety of social science disciplines. During this process, the researchers assumed that the participants' responses were sincere and correct, and data analysis was carried out based on that strong assumption (Moors, 2010). However, whether respondents' responses accurately reflect their sincere opinions has never been fully addressed. If this assumption is not met, researchers may obtain results that prevent them from drawing valid conclusions. Even in the earliest psychometrics literature, it was stated that participant insincerity jeopardizes the validity of inferences obtained from survey data (Cronbach, 1946, 1950). In addition, when the scores on the estimated trait do not precisely reflect the true level of the trait being evaluated, this is referred to as a measurement error (Lord & Novick, 1968). For this reason, it is critical to identify participants whose responses do not accurately reflect their genuine opinions on the topics that the measurement tool intends to measure and to minimize measurement error by eliminating these individuals' responses from the dataset. Only in this way, a tester can draw valid inferences based on the measurement process.

Traditionally, the degree of measurement error is examined at the item level and indicates how well an item's observed response pattern fits into the proposed test model (Lomax, 1986). If the pattern does not contribute to the hypothesized model, it is accepted that at least one item does not fit well into the measurement model, and these item(s) are often removed from the measurement tool or revision of these item(s) is considered.

Persons, on the other hand, can be another source of error (Viswanathan, 2005). In general, error-producing responses at the person level might be caused by carelessness, a lack of desire, a specific response style, deceptive responses, and so on. These test-taker behaviors induce respondents to produce inconsistent and unexpected (sometimes referred to as aberrant) response patterns. The emergence of aberrant response patterns jeopardizes the reliability and validity of the measurement process (Reynolds, 2010). When the sample size is large enough, and the ratio of individuals with aberrant response patterns to the sample group is low, the renewal of the analysis by removing aberrant response patterns from the sample group provides better results in terms of model fit (Ferrando, 2007).

Evaluating the observed response patterns on a personal level can be examined by investigating "person-fit" or "fit at the person-level" (Meijer & Sijtsma, 2001). There are several reasons why it is critical to examine person-fit. However, the fundamental explanation is that the observed scores of persons with a lower degree of fit may not

accurately reflect their underlying ability levels (Meijer, Niessen, & Tendeiro, 2016). In this case, conclusions drawn from observed scores may be invalid. As a result, observed response patterns of persons must be routinely checked for possible aberrancy (International Test Commission, 2014). In addition, person-fit statistics are useful indicators of response consistency/validity index (Reise & Waller, 1990).

Many techniques for assessing person-fit have been developed over time, including latent class and mixture analysis, diagnostic classification modeling, multilevel logistic regression analysis, factor analysis/covariance structure modeling (Rupp, 2013), and a recently proposed family of techniques known as “person-fit statistics”. The focus of the person-fit statistics is on the observed response vectors of respondents. They are used to determine how well an individual's item response vector matches the expected response pattern from a given measurement model. In fact, numerous person-fit indexes have been developed to date to assess the validity of the individual response pattern within the IRT framework. Even though they employ different computational procedures (e.g., Tatsuoka, 1996; Drasgow, Levine, & McLaughlin, 1987), the overarching goal of all is to assess the degree of consistency of a response pattern with a specified IRT model (Embretson & Reise, 2000).

Due to the importance of investigating model fit at the person level, several researchers have begun to focus on the topic of person-fit statistics in recent years (Ferrando & Chico, 2003; Meijer & Sijtsma, 2001; Pan & Yin, 2017). While some research in the person-fit literature compares different person-fit statistics in terms of power and type I error (e.g., Karabatsos, 2003), others studied the causes and potential impacts of person misfit on test features (e.g. Petridou & Williams, 2007). On the other hand, the existing person-fit statistics are unable to identify the probable causes of the misfit. Examining the links between person-fit and certain key variables, on the other hand, is the only way to better understand how person misfit occurs. In line with this reality, recent studies focus on how test and personality characteristics are related to person-fit. A recent study (Dodeen & Darabi, 2009) reported that personality traits are related to aberrant responses. For example, a test taker with particular personality traits may try to give a better impression when answering a personality instrument and choose response options with higher or lower scores (Zickar & Drasgow, 1996). Another recent study dealt with the question of whether or not there is an association between being a misfit and outlier (Felt, Castaneda, Tiemensma & Depaoli, 2017) and concluded that misfit respondents are probably an outlier. In addition,

the researchers stated that person-fit statistics could be an effective diagnostic tool to detect outliers. Thus, another probable cause for respondents' endorsement of extreme response alternatives. All in all, the relevant literature support that misfit respondents may be using extreme response more.

In the relevant literature, this kind of tendency of an individual to respond to test items in a specific and systematic manner on a basis different from the trait that the items are intended to measure is called response style (Paulhus 1991). Weijters (2006) divided response style sources into two main categories: stimulus level and respondent level. At the stimulus level, the response style is seen as a result of the measurement tool, while at the respondent level, the response style is seen as a result of personal characteristics. Examples of possible sources of response style at the stimulus level are the item format (Weijters, Cabooter & Schillewaert, 2010) and the mode of data collection. At the respondent level, the psychological and demographic characteristics of the individual are the main sources.

Response styles, in general, contribute to systematic error. The univariate and multivariate distributions of gathered data result in invalid and erroneous response patterns. Nonetheless, researchers frequently fail to include an analysis of the ERS in their studies. Readers who are interested in gaining more information about these response styles are recommended to look at relevant literature (i.e., Baumgartner & Steenkamp, 2001; Hurley, 1998).

The most known is the ERS. It corresponds to the disproportionate preference of endpoints in Likert type or rating scales, regardless of the content of the items. It is also possible to consider the ERS as the opposite of endorsing middle options among possible response categories. For example, the ERS for a five-point Likert scale item relates to the number of items for which a participant selects “strongly agrees with” or “strongly disagrees with” options (Greenleaf, 1992). The traditional method of calculating the ERS amount is used by simply adding the extreme replies of a respondent to a range of items (Baumgartner & Steenkamp 2001). It is also known as the “counting technique”. Vaerenbergh and Thomas (2013) provided a detailed review of alternative techniques to detect the ERS.

Persons using the ERS are likely to produce response patterns that the trait level of the respondent cannot estimate. In addition, the disproportionate use of extreme responses alters the correlations between the variables and makes it difficult to understand the nature of individual differences. Many different analyzes such as calculation of internal consistency, regression analysis, factor analysis, and structural equation models are based on

correlations (Baumgartner & Steenkamp, 2001). Therefore, ignoring the effect of the extreme responses may lead to misleading results. Despite its importance and the effect of measurement results, controlling person-fit is not frequent in real-world testing applications (Meijer, Niessen & Tendeiro, 2016).

In addition, the connection between the use of the ERS and personality is implied in the literature. Many studies have shown that the ERS use of individuals remains relatively stable over time (Jain & Agrawal, 1977; Greenleaf, 1992). In addition, the previous studies revealed that the use of the ERS is highly prevalent when using self-report measurement tools. A recent study reported that 25% of the participants adopted the ERS during the response process (Austin, Deary & Egan, 2006). The consistency and stability of these response styles (e.g., Weijters, Geuens & Schillewaert, 2010) implies that individuals' traits and even personalities may impact using the ERS. According to the previous findings, there was a positive relationship between conscientiousness and extreme reaction styles. In addition, it was found that extrovertedness is related to the ERS (Austin et al., 2006; Naemi, Beal & Payne, 2009). The findings were also confirmed by He and van de Vijver (2013). Finally, in another study conducted by Hamilton (1968), it was stated that there was a negative relationship between neuroticism and ERS (He & van de Vijver, 2013). All in all, consistent results have been reached on the relationship between personality and the ERS (Kyllonen, 2008; Landers, Sackett & Tuzinski, 2011)

As previously stated, evaluating response patterns is critical for identifying the source of measurement error. Because item response theory (IRT) is also a model-based measurement approach, detecting misfit individuals is especially important within the IRT framework. In addition, earlier studies frequently investigated person-fit when evaluating achievement exams (Brown & Villareal, 2007). However, like other approaches in the field of psychometrics, person-fit statistics have come to be employed in psychological assessment procedures where instruments contain polytomous items (Dodeen & Darabi, 2009; Ferrando, 2012). The ERS can have a detrimental impact on the person-fit statistics because it will generate response patterns incompatible with the measurement model utilized. In line with this possibility, the purpose of this study was to investigate whether there is an association between person-fit and the ERS.

Method

This research, which aims to examine the association between person-fit and extreme response style, is relational research since it aimed to determine the connection between these two variables.

Participants

Due to the 2020 pandemic outbreak, participation in the study was held online. Because it is not easy to apply systematic sampling procedures when using an online data collection approach, the data was collected using convenient sampling. The participants were selected from four different state universities in a metropolitan province and included 727 students from 11 different departments. The demographic characteristics of the sample group are provided in Table 1 below.

Accordingly, the participants consisted of 590 females ranging in age from 16 to 23 ($M=21.19$, $SD=1.70$) and 137 males ranging in age from 17 to 23 ($M=21.39$, $SD=1.47$). Regarding departments, the participants contained History ($n=32$), Psychology ($n=120$), Dentistry ($n=104$), Child Development ($n=69$), Psychological Counseling ($n=65$), Science Teaching ($n=25$), Math Teaching ($n=102$), Social Sciences Teaching ($n=42$), Literature ($n=69$), Sociology ($n=8$), Fine Arts ($n=97$) students. This final sample size was obtained after eliminating 73 participants because of missing data and being an outlier. It was scrutinized whether or not the removal of these 73 participants disturbs the characteristics of the sample group in terms of age, gender, and departments they enrolled and found that it was not the case.

Table 1. *Descriptive statistics for the demographical characteristics of the participants*

Variable	Groups	n	%
Gender	Female	590	81.2
	Male	137	18.8
Department	History	26	3.6
	Psychology	120	16.5
	Dentistry	104	14.3
	Child Development	69	9.5
	Psychological Counseling	65	8.9
	Science Teaching	25	3.4

Math Teaching	102	14.0
Social Sciences		
Teaching	42	5.8
Literature	69	9.5
Sociology	8	1.1
Fine Arts	97	13.3

Data Collection Tools

When the literature presented in the introduction is reconsidered as a whole, it is fair to assume that as the respondents answer to a personality inventory, participants will possibly use the ERS or become a misfit. For this reason, a personality inventory was preferred to investigate an association between the ERS and the person-fit. Accordingly, the Basic Personality Traits Inventory (BPTI) was selected as a data collection instrument. The BPTI is a self-report instrument developed by Gencöz and Öncül (2012) for University Students. It is a five-point Likert-type scale consisting of six dimensions measured by 45 adjective items. The participants were asked to what extent these adjectives represent themselves. In addition to conventional five personality dimensions, including five dimensions (Extraversion-E, Conscientiousness-C, Agreeableness-A, Neuroticism-N, and Openness-O), the scale also includes a sixth one: Negative Valence-NV. During the development phase, exploratory factor analysis was carried out to investigate the factorial structure of the BPTI, and the results revealed that the six dimensions explained 53% of total variance and items loaded on the factors with loadings ranging between .84 and .40. BPTI was also evaluated in terms of internal consistency and test-retest reliability by Gencöz and Öncül, and it was found that internal consistency coefficients were ranged between .71 and .89 for sub-dimensions. Test-retest reliability values were also found to vary between .71 and .84. These coefficient values supported the acceptable level of internal consistency for the BPTI items. For the current study, contrary to the conventional approach, the internal consistency of the BPTI items was not calculated because Person-fit was investigated in the IRT context, which approaches the concept of reliability differently.

Data Analysis

When the person-fit literature is examined, it was seen that there are many different statistics developed so far. It is recommended that interested readers should look at Meijer and

Sijtsma (2001) for a review of these methods. The lz person-fit statistics (Drasgow, Levine & Williams, 1985) were preferred in the current study because it is the most studied and the most known one. In addition, it is based on a strong psychometric approach, also known as item response theory. Once the data set was obtained, the lz statistics were calculated for each participant.

The lz statistic is the standardized log-likelihood of response vectors of individuals and has a similar distribution with standardized z-values, and high negative values are interpreted as a misfit. When the lz statistics are calculated, it is necessary to set a threshold value based on which a decision is made whether a person is regarded fit or misfit. For the current study, the bootstrap method was used with $\alpha = 0.05$ to determine the threshold value.

After the lz statistics were calculated and threshold values were determined, a new dichotomous variable was created based on the lz values where 0 represents misfit persons and 1 represents fitting ones. In addition, two other dichotomous variables were also created to specify lower ERS (showing whether the person selects the lowest response category) and upper ERS (showing whether the person selects the highest response category) usage for each individual and for each item of the BPTI. If a person shows the lower ERS or the upper ERS, it is accepted that this person is using the ERS. Finally, for each question of the BPTI, the chi-square statistics were calculated for the dichotomous ERS variable and the dichotomous person-fit variables. In this way, it was aimed to see the association between the ERS and person-fit. The calculation of lz statistics was performed using the “PerFit” package (v0.3.4; Tendeiro, Meijer & Niessen, 2016) included in the R statistical program (v4.0.0; R Core Team, 2019). SPSS, version 21, was used for the other statistical procedures.

Findings

The Pearson correlation coefficient values between the dimensions of BPTI, the mean, standard deviation, skewness, kurtosis values of the scores for each dimension were provided in Table 2 below. The correlation values between all dimensions were found to be significant ($p < 0.01$). In addition, when the kurtosis and skewness values were analyzed, it was seen that the distributions of those values were in a range of +1 and -1, which implies that the distribution of sub-dimension scores can be accepted as normal.

Table 2. Descriptive Statistics of and Correlations Between Personality Types

	Dimensions	1	2	3	4	5	6
1	Extraversion		0.28**	0.30**	-0.21**	0.58**	-0.22**
2	Conscientiousness			0.32**	-0.22**	0.40**	-0.23**
3	Agreeableness				-0.17**	0.38**	-0.40**
4	Neuroticism					-0.17**	0.34**
5	Openness						-0.15**
6	Negative Valence						
	M=	18.72	22.35	33.57	27.81	29.39	34.15
	SD=	4.92	6.06	8.34	6.45	5.73	4.14
	Skewness=	-0.10	-0.38	-0.16	-0.20	-0.54	-0.54
	Kurtosis=	-0.01	-0.40	-0.19	-0.56	0.09	0.07

** $p < 0.01$

The descriptive statistics of the lz values, cut-point values to decide whether or not a participant will be coded as fit or not fit, the frequency statistics showing the number of participants coded as fit, and as a misfit for each sub-dimension were provided in Table 3 below. Notably, the highest number of misfit persons were observed for the conscientiousness (n=66) sub-dimension, while the lowest number of misfit persons (n=49) was observed for negative valence. In addition, the connection between the correlation among the dimensions and similarity of the number of misfit persons was investigated qualitatively. It was observed that any kind of conclusive result could be inferred from this prior investigation.

In order to examine the relationship between the ERS and person-fit, dichotomized variables (using ERS vs. not using ERS & being fit vs. being not fit) were used to calculate Chi-square statistics for each item in the BPTI. The ERS usage was dichotomized so that if a person uses the lowest response category (named as lower ERS) or the highest response category (named as upper ERS), that person was coded as ERS user. The results are

provided in Table 4 below. Based on findings, it was realized that for 24 items, there was a dependency of being a misfit and using the ERS. That is to say, being fit or misfit is significantly dependent on using the ERS for those items.

Table 3. *Descriptive statistics for estimated person-fit statistics*

	Min.	Max.	M	SD	Skewness	Kurtosis	Cut P.	Fit #	Misfit #
Extraversion	-4.74	1.89	0.27	1.10	-1.26	2.12	-1.48	670	57
Conscientiousness	-7.31	1.97	0.25	1.19	-1.60	3.56	-1.42	671	66
Agreeableness	-9.89	1.64	0.18	1.20	-2.54	12.20	-1.50	674	53
Neuroticism	-5.55	2.25	0.30	1.10	-1.30	2.80	-1.36	667	60
Openness	-5.52	1.50	0.27	1.03	-1.52	2.80	-1.52	677	50
Negative Valence	-3.62	1.38	0.10	0.75	-1.45	2.73	-1.12	678	49

This finding tells that ERS usage frequency among fit and misfit participants shows the difference. When the table is examined in more detail, it was seen that misfit individuals use both lower and upper ERS more frequently than fit persons for 17 of 24 items. On the other hand, lower ERS was used in 5 items for the fit participants, while upper ERS was used more for only two items. In short, it can be said that misfit individuals more intensely use the ERS.

Table 4. *Item level comparison of response styles by person-fit groups*

Items	ERS (Lower)			ERS (Upper)			<i>p</i>
	Misfit	Fit	Diff.	Misfit	Fit	Diff.	
Extraversion							
Timid	10.5	6.1	4.4	17.5	14	3.5	
Withdrawn	68.4	70.4	-2	0	0.1	-0.1	
Shy	0	0.6	-0.6	31.6	35.1	-3.5	
Talkative	10.5	3	7.5	35.1	21.8	13.3	**
Lethargic	1.8	1.5	0.3	19.3	20.9	-1.6	
Enterprising	15.8	4.2	11.6	24.6	27.9	-3.3	**
Cold	17.5	8.5	9	19.3	11.3	8	*
Passive	3.5	0.7	2.8	47.4	41.6	5.8	
Conscientiousness							
Self-disciplined	18.2	14.4	3.8	21.2	11.5	9.7	

Tidy	4.5	2.3	2.2	18.2	14.2	4	
Hard-working	22.7	30	-7.3	15.2	3.8	11.4	**
Prudent	10.6	3.2	7.4	28.8	14.7	14.1	**
Fussy	60.6	66	-5.4	4.5	1.8	2.7	**
Determined	9.1	4.7	4.4	19.7	14.8	4.9	
Irresponsible	4.5	0	4.5	65.2	49.5	15.7	**
Lazy	1.5	0.2	1.3	68.2	47.7	20.5	**
Agreeableness							
Sincere	0	1.6	-1.6	32.1	26.1	6	
Compassionate	13.2	20.8	-7.6	17	5.8	11.2	*
Genial	3.8	0	3.8	45.3	45.8	-0.5	**
Well intentioned	5.7	1.9	3.8	37.7	21.5	16.2	**
Philanthropic	15.1	23.1	-8	9.4	8.3	1.1	
Tolerant	9.4	2.4	7	50.9	46.3	4.6	*
Sharer	7.5	0.6	6.9	37.7	37.2	0.5	**
Sensitive	3.8	1.2	2.6	34	38.4	-4.4	
Neuroticism							
Nervous	10	4.5	5.5	21.7	21.1	0.6	
Aggressive	61.7	62.4	-0.7	3.3	1.3	2	
Angry	31.7	15.1	16.6	35	7.8	27.2	**
Temperamental	1.7	0.3	1.4	43.3	34.8	8.5	
Impatient	35	16.8	18.2	25	5.8	19.2	**
Capricious	48.3	31.9	16.4	11.7	3.6	8.1	**
Impetuous	80	81.7	-1.7	0	0.1	-0.1	
Touchy	6.7	4.6	2.1	35	16.8	18.2	**
Worried	23.3	11.5	11.8	16.7	10.5	6.2	*
Openness to Experience							
Self-confident	86	80.8	5.2	4	0.6	3.4	*
Self-assured	8	13.9	-5.9	30	16.2	13.8	*
Brave	12	2.4	9.6	28	18.2	9.8	**
Creative	36	37.5	-1.5	12	3.4	8.6	*
Easy going	12	5.8	6.2	22	27	-5	
Capable	6	6.2	-0.2	30	22.3	7.7	

Negative Valence

Ill-mannered	12	17.7	-5.7	14	9.7	4.3	
Pretentious	2	0.4	1.6	42	39.7	2.3	**
Rude	6	3.4	2.6	40	28.8	11.2	
Backstabbing	8	3.1	4.9	36	27.9	8.1	
Greedy	4	1.8	2.2	42	30.1	11.9	
Hidebound	8	1.5	6.5	42	30.7	11.3	**

Note: ** $p < .05$; * $p < .01$; ERS (lower)=percentage of participants endorsing highest lowest category; ERS (Upper)= percentage of participants endorsing highest category; Diff.= difference of percentage values between fit and misfit group; All values in the table show percentages.

To indicate the relationship between person-fit and use of the ERS, a small portion of response matrices for persons with the highest and lowest fit (ones who have lowest and highest lz fit statistic values) were presented in Table 5. The table also contains the average upper and lower ERS use by both groups. The results show that ten individuals with the highest misfit level used an average of 1.7 times lower ERS in 8 items, while the average use of ERS (lower) was 2.6 for the 10 for those persons. For the participants having the highest fit level, the number or lower ERS was 0.9, while this average number was 1.1. for upper ERS use. That is to say, misfit persons are more differentiated for lower and upper ERS use. They generally use upper ERS more frequently. The Chi-square test was used to test whether being fit or not fit is related to using ERS in the upper or lower tail. The result provided no proof for this relationship ($p > .05$). That is to say, regardless of which kind of ERS a respondent use, misfit persons use more ERS than fit ones. This finding showed that the use of the ERS is more intense for individuals with the lowest lz values (highest misfit participants). This finding was consistent for both upper and lower ERS use.

Table 5. Response matrices and average use of ERS for a person with highest and lowest lz values

	i1	i2	i3	i4	i5	i6	i7	i8	ER (lower) #	ER (upper) #
Misfit Person 1	5	5	5	1	1	5	5	5	2	6
Misfit Person 2	5	5	5	4	5	5	5	5	0	7
Misfit Person 3	1	3	3	3	1	3	5	5	2	2
Misfit Person 4	5	5	3	1	1	3	5	5	2	4

Misfit Person 5	1	4	3	3	1	4	1	1	4	0	
Misfit Person 6	2	2	4	2	2	1	2	4	1	0	
Misfit Person 7	5	4	3	5	1	2	4	5	1	5	
Misfit Person 8	4	3	1	4	4	4	4	4	1	0	
Misfit Person 9	3	4	1	2	1	4	5	5	2	2	
Misfit Person 10	3	4	4	1	1	4	2	3	2	0	
									Mean	1.7	2.6
Fit Person 1	3	4	4	4	1	4	5	5	1	2	
Fit Person 2	3	5	3	4	1	4	5	5	1	3	
Fit Person 3	4	2	2	4	2	3	4	4	0	2	
Fit Person 4	4	5	3	4	3	4	5	4	0	2	
Fit Person 5	2	4	1	4	1	4	4	4	2	0	
Fit Person 6	3	4	4	4	2	3	4	4	0	0	
Fit Person 7	2	3	3	4	1	3	4	4	1	0	
Fit Person 8	4	5	3	4	1	5	4	5	1	2	
Fit Person 9	2	3	3	3	1	3	3	3	1	0	
Fit Person 10	1	4	3	3	1	2	3	3	2	0	
									Mean	0.9	1.1

Discussion

As mentioned in the introduction, people's response styles cause them to elicit scores that do not indicate the trait level they possess. Because item response theory employs a model-based approach to measurement, persons' estimated ability levels will likely become incompatible with the model if other processes influence response patterns. In this case, these individuals do not fit the model specified. In this study, how the response style is related to person-fit was investigated. Accordingly, calculated lz statistics of the individuals showing the level of fit were re-coded into two categories (fit and misfit). In addition, the response patterns of the individuals were examined and the variables specifying whether lower or upper ERS was used created for each item. Afterward, the relationships between these categorical variables were examined with chi-square statistics for each item. As a result, for most of the items, a statistically significant relationship was found between person-fit and the ERS. A similar research question was investigated in simulation studies and obtained similar and comparable results (Tendeiro, 2017; Adams et al., 2019). Unlike

this previous study, the current study used real data and further supported the connection between person-fit and the ERS.

In addition, the response patterns of the ten individuals with the highest and the lowest I_z values were examined and compared according to their use of the ERS. Similarly, it has been confirmed that misfit persons use more ERS. When the findings are examined in more detail, it can be said that the most consistent findings are observed for the items in the sub-dimension of neuroticism. When the relationship between the ERS and person-fit was examined for this dimension, it was observed that both lower ERS and upper ERS were observed more for the misfit persons on all items for this dimension. This main result is in line with the previous literature (see Felt et al., 2017), where the connection between outliers and person fit was investigated. Even outliers do not necessarily use extreme responses; being an outlier and using extreme responses are highly overlapping phenomena.

Conclusions

In this study, the results should be read carefully because only one measurement tool was used to determine aberrant responses of individuals, and only one person-fit statistic was used to determine misfit persons. It was suggested that readers look at St-Onge et al. (2011), to illustrate the negative effects of employing only one person-fit statistic. As stated in the introduction section, there are many different fit statistics in the literature. Bu using those alternative fit statistics and multiple personality scales simultaneously in one study, the results' validity will improve.

Due to the relatively few studies on this subject, the present study has not been conducted and discussed based on other preliminary studies. As a result, readers can develop alternative explanations and interpretations for the findings. As a result, the conclusions reached in this study should be expanded. It can be said that the findings obtained in the current study need to be verified or refuted by future studies. In addition, only the ERS, among other response styles, was considered in this study. As stated in the introduction, there are many different response styles, and examining how each of these response styles related to person-fit will contribute to the existing findings. Finally, as a result of working with real data, not all processes affecting participants' responses could be controlled. Therefore, evaluating the observed response patterns in the context of ERS may provide deceptive outcomes. For example, it has not been checked whether the ERS use is caused by careless responses or by participants' desire to present themselves as more socially desirable. Therefore, it is recommended to interpret the results carefully.

Although personality data was used in the current study, the results can draw limited pedagogical inferences because psychological variables are closely related to their educational outcomes, and the relationship between personality and academic achievement was already reported in the literature (Nambudiri, Shaik & Ghulyani, 2019). Not removing misfit individuals from the data set may lead to erroneous diagnoses and misleading results prevent educators from taking proactive actions to foster educational outcomes of students. It is clear that educational outcomes can not be fostered without valid and reliable measurement of psychological variables

Karabatsos (2003) stated that an aberrant response rate of 25% and above is the threshold in terms of acceptability. In this study, the proportion of individuals with misfit was found to be below 10%. Therefore, under current conditions where the level of person-fit proportion is well below the acceptable threshold, examining person-fit in the context of any variable may not be healthy to draw definitive results.

The effect of a third variable can be examined in future studies. For example, examining the question of how the relationship between ERS and person-fit changes according to the gender variable will enrich the existing knowledge on this subject. In addition, it will be useful to examine how the relationship between person-fit and the ERS changes according to the scores obtained for each personality sub-dimensions. In this way, the effect of personality traits of individuals could be linked to the current findings. Finally, it is recommended to examine the model fit levels of the items ERS use differs according to the person-fit. It was believed that examining how the ERS affects IRT-based item fit in the context of person-fit will provide a great contribution to the current literature.

Ethical Approval: *This research was conducted with the permission of the Marmara University Institute of Educational Sciences Research and Publication ethics committee with decision no 2020.8.1.82 dated 19.11.2020*

Conflict Interest: *The author declares no conflict of interest*

References

- Adams, D. J., Bolt, D. M., Deng, S., Smith, S. S., & Baker, T. B. (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of Mathematical and Statistical Psychology*, 72(3), 466-485.
- 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(6), 1235-1245.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 143-156.
- Brown, R. S., & Villarreal, J. C. (2007). Correcting for person misfit in aggregated score reporting. *International Journal of Testing*, 7(1), 1-25.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475-494.
- Cronbach, L. J. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement*, 10(1), 3-31.
- Dodeen, H., & Darabi, M. (2009). Person-fit: relationship with four personality tests in mathematics. *Research Papers in Education*, 24(1), 115-126.
- Drasgow, F., Levine, M. V., and Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38(1), 67-86.
- Embretson, S. E., Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates.
- Felt, J. M., Castaneda, R., Tiemensma, J., and Depaoli, S. (2017). Using Person Fit Statistics to Detect Outliers in Survey Research. *Frontiers in psychology*, 8, 863.
- Ferrando, P. J. (2007). Factor-analytic procedures for assessing response pattern scalability. *Multivariate Behavioral Research*, 42(3), 481-507.
- Ferrando, P. J. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. *Personality and Individual Differences*, 52(6), 718-722.

- Ferrando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement, 61*(6), 997-1012.
- Gençöz, T. & Öncül, Ö. (2012). Examination of Personality Characteristics in a Turkish Sample: Development of the Basic Personality Traits Inventory. *The Journal of General Psychology, 139*(3), 194-216.
- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly, 56*(3), 328-351.
- Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology, 132*(2), 201-210.
- International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. *International Journal of Testing, 14*(3), 195-217.
- Jain, U., & Agrawal, L. (1977). Generality of extreme response style. *Journal of psychological researches, 21*, 67-72.
- Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. M. (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software, 74*(1), 1-27.
- Tendeiro, J. N. (2017). The lz (p)* Person-fit statistic in an unfolding model context. *Applied Psychological Measurement, 41*(1), 44-59.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education, 16*(4), 277-298.
- Kyllonen, P. C. (2008). *The research behind the ETS Personal Potential Index (PPI)*. Princeton, NJ: Educational Testing Service.
- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology, 96*(1), 202.
- Lomax, R. G. (1986). The effect of measurement error in structural equation modeling. *The Journal of Experimental Education, 54*(3), 157-162.
- Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley.

- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107-135.
- Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2016). A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program. *Assessment*, 23(1), 52-62.
- 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(1), 93-119.
- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of personality*, 77(1), 261-286.
- Nambudiri, R., Shaik, R., & Ghulyani, S. (2019). Student personality and academic achievement: mediating role of psychological capital (PsyCap). *International Journal of Educational Management*. 4(34): 767-781.
- Paulhus, D. L. (1991), "Measurement and control of response bias," in *measures of personality and social psychological attitudes*, John P. Robinson, Phillip R. Shaver, and Lawrence S. Wright, eds. San Diego: Academic Press, 17-59.
- Petridou, A., & Williams, J. (2007). Accounting for aberrant test response patterns using multilevel models. *Journal of Educational Measurement*, 44(3), 227-247.
- Reise, S. P., and Waller, N. G. (1990). Applications of response pattern aberrancy analysis to personality assessment. *Unpublished manuscript*.
- Reynolds, C. R. (2010). Measurement and assessment: An editorial view. *Psychological Assessment*, 22, 1-4.
- Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, 55(1), 3.
- St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2011). Accuracy of person-fit statistics: A Monte Carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, 35(6), 419-432.
- Tatasuoka, K. (1996). Use of generalized person-fit indexes, zetas for statistical pattern classification. *Applied measurement in education*, 9(1), 65-75.

- Pan, T., & Yin, Y. (2017). Using the Bayes factors to evaluate person fit in the item response theory. *Applied Measurement in Education, 30*(3), 213-227.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International journal of public opinion research, 25*(2), 195-217.
- Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks: Sage Publications.
- Weijters, B. (2006). *Response styles in consumer research*. Doctoral Dissertation. Ghent University. Faculty of Economics and Business Administration, Ghent, Belgium.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236-247.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological methods, 15*(1), 96.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied psychological measurement, 20*(1), 71-87.



Kişi Uyumu ve Uç Yanıt Stili Arasındaki İlişki

Akif AVCU*

• **Geliş Tarihi:** 06.12.2020 • **Kabul Tarihi:** 24.11.2021 • **Çevrimiçi Yayın Tarihi:** 24.11.2021

Öz

Uç yanıt stili (UYS), içeriği ne olursa olsun, Likert tipi veya derecelendirme ölçeği ile puanlanan maddelerde uç yanıt seçeneklerinin orantısız bir şekilde tercih edilmesini ifade eder. UYS, kullanılan ölçüm modelinin uyumunu bozma potansiyeline sahiptir. Mevcut çalışmada, UYS'nin kullanımının normal olmayan yanıt modellerinin olası bir nedeni olduğu varsayılmıştır. Madde tepki kuramı (MKT) bağlamında, normal olmayan yanıt örüntüleri aynı zamanda uyum göstermeyen yanıtlar olarak da bilinir ve bu örüntüler, kişi-uyum istatistikleri kullanılarak incelenir. Buna doğrultuda, bu çalışmanın amacı, UYS'nin uyum göstermeyen yanıt örüntüleri ile bağlantılı olup olmadığını araştırmak olarak belirlenmiştir. Bu amaç doğrultusunda uygun örnekleme yöntemi kullanılarak Temel Kişilik Özellikleri Envanteri (TKÖE) ile 727 üniversite öğrencisinden veriler toplanmıştır. Elde edilen bulgular, UYS'yi kullanan kişilerin uyum göstermeyen yanıt örüntülerini üretme olasılıklarının daha yüksek olduğunu göstermiştir. Elde edilen sonuçlar, hem alt uç hem de üst uç UYS kullanımı için tutarlıdır. Bulgulara dayanarak, UYS'nin kişilik ölçümü bağlamında kişi uyum göstermemesinin muhtemel bir sebebi olduğu sonucuna varılmıştır. UYS ve kişi uyumu arasındaki bağlantının daha derinlemesine analiz edilmesi önerilmiştir.

Anahtar sözcükler: uç yanıt stili, kişi uyumu, madde tepki kuramı.

Atıf:

Avcu, A. (2022). Kişi uyumu ve uç yanıt stili arasındaki ilişki. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 55, 27-45. doi:10.9779.pauefd.836567

* Arş Gör Dr, Marmara Üniversitesi, Atatürk Eğitim Fakültesi, ORCID: 0000-0003-1977-7592, e posta: avcuakif@gmail.com

Giriş

Öz bildirim ölçüm araçları, sosyal bilimlerin farklı disiplinlerinde veri toplama yöntemi olarak yaygın olarak kullanılmaktadır. Bu süreçte araştırmacılar, katılımcıların yanıtlarının samimi ve doğru olduğunu varsayar ve bu güçlü varsayıma dayalı olarak veri analizi gerçekleştirilir (Moors, 2010). Ancak, katılımcıların yanıtlarının samimi görüşlerini doğru bir şekilde yansıtmayı yansıtmadığı sorusunun üzerinde hiçbir zaman durulmaz. Eğer bu varsayım karşılanmazsa, araştırmacılar geçerli çıkarımlara varmalarını engelleyen sonuçlar elde edebilir. En eski psikometri alanyazınında bile, katılımcı samimiyetsizliğinin anket verilerinden elde edilen çıkarımların geçerliliğini tehlikeye attığı belirtilmiştir (Cronbach, 1946, 1950). Ayrıca, kestirilen özelliğin puanları değerlendirilen özelliğin gerçek seviyesini tam olarak yansıtmadığında bu durum ölçüm hatası olarak adlandırılır (Lord ve Novick, 1968). Bu nedenle, ölçme aracının ölçmeyi amaçladığı konularda gerçek görüşlerini doğru yansıtmayan katılımcıları tespit etmek ve bu kişilerin yanıtlarını veri setinden çıkararak ölçme hatasını en aza indirmek kritik öneme sahiptir. Ancak bu şekilde, testi uygulayan kişi ölçüm sürecine dayalı olarak geçerli çıkarımlar yapabilir.

Geleneksel olarak, ölçme hatasının derecesi madde düzeyinde incelenir ve bir maddenin gözlenen yanıt örüntüsünün önerilen test modeline ne kadar iyi uyduğunun göstergesidir (Lomax, 1986). Örüntü, varsayılan modele katkıda bulunmuyorsa, en az bir maddenin ölçme modeline tam olarak uymadığı kabul edilir ve bu madde(ler) genellikle ölçme aracından çıkarılır veya bu madde(ler)in revizyonu gerçekleştirilir.

Diğer taraftan, kişiler de hatanın bir diğer kaynağı olabilir (Viswanathan, 2005). Genel olarak, kişi düzeyinde hata üreten yanıtlara dikkatsizlik, isteksizlik, belirli bir tepki stili, aldatıcı tepkiler vb. neden olabilir. Bu davranışlar, yanıt verenlerin tutarsız ve beklenmedik (bazen anormal olarak da anılan) yanıt örüntüleri üretmesine neden olur. Anormal tepki örüntülerinin ortaya çıkması, ölçüm sürecinin güvenilirliğini ve geçerliliğini de tehlikeye atmaktadır (Reynolds, 2010). Örneklem büyüklüğü yeterince büyük olduğunda ve anormal tepki örüntülerine sahip bireylerin örneklem grubuna oranı düşük olduğunda, örneklem grubundan anormal tepki örüntülerinin çıkarılarak analizin yenilenmesi model uyumu açısından daha iyi sonuçlar vermektedir (Ferrando, 2007).

Kişi düzeyinde gözlenen yanıt örüntüleri, “kişi-uyum” veya “kişi düzeyinde uyum” araştırılarak incelenebilir (Meijer ve Sijtsma, 2001). Kişi uyumunu incelemenin önemli olmasının birkaç nedeni vardır. Bununla birlikte, temel açıklama, daha düşük uyum derecesine sahip kişilerin gözlemlenen puanlarının, altta yatan yetenek seviyelerini doğru bir

şekilde yansıtmayabileceğidir (Meijer, Niessen ve Tendeiro, 2016). Bu durumda, gözlemlenen puanlardan çıkarılan sonuçlar geçUYSiz olabilir. Sonuç olarak, kişilerin gözlemlenen tepki örüntüleri olası anormalliklere karşı rutin olarak kontrol edilmelidir (Uluslararası Test Komisyonu, 2014). Ayrıca, kişi-uyum istatistikleri, yanıt tutarlılığı/geçerlilik indeksinin kullanışlı göstergeleridir (Reise ve Waller, 1990).

dikkate alınan.

Gizli sınıf ve karışım analizi, tanısal sınıflandırma modellemesi, çok düzeyli lojistik regresyon analizi, faktör analizi/kovaryans yapısı modellemesi (Rupp, 2013) ve yakın zamanda önerilen ve kişi uyumunu değerlendirmek üzere kullanılan kişi uyum istatistikleri ailesi kişi uyumunu değerlendirmek için zaman içinde geliştirilmiş olan birçok teknik arasındadır. Kişi-uyum istatistiklerinin odak noktası, yanıt verenlerin gözlemlenen yanıt vektörleridir. Bir bireyin madde yanıt vektörünün belirli bir ölçüm modelinden beklenen yanıt modeliyle ne kadar iyi eşleştiğini belirlemek için kullanılırlar. Aslında, MTK çerçevesinde bireysel yanıt modelinin geçerliliğini değerlendirmek için bugüne kadar çok sayıda kişi uyum indeksi geliştirilmiştir. Farklı hesaplama işlemleri kullanmalarına rağmen (örneğin, Tatsuoka, 1996; Drasgow, Levine ve McLaughlin, 1987), hepsinin genel amacı, belirli bir MTK modeli ile bir yanıt örüntüsünün tutarlılık derecesini değerlendirmektir (Embretson ve Reise, 2000).

Kişi düzeyinde model uyumunun araştırılmasının önemi nedeniyle, son yıllarda birçok araştırmacı kişi-uyum istatistikleri konusuna odaklanmaya başlamıştır (Ferrando ve Chico, 2003; Meijer ve Sijtsma, 2001; Pan ve Yin, 2017). Kişi-uyumu ile ilgili alanyazındaki bazı araştırmalar, farklı kişi-uyum istatistiklerini güç ve birinci tip hata açısından karşılaştırırken (örn., Karabatsos, 2003), diğerleri, kişi uyumsuzluğunun test özellikleri üzerindeki etkilerini ve potansiyel nedenlerini incelemiştir (örn. Petridou ve Williams, 2007). Öte yandan, mevcut kişi-uyum istatistikleri, uyumsuzluğun olası nedenlerini belirleyememektedir. Öte yandan, kişi uyumsuzluğu ile belirli anahtar değişkenler arasındaki bağlantıları incelemek, kişi uyumsuzluğunun nasıl oluştuğunu daha iyi anlamının tek yoludur. Bu gerçeğe örtüşecek şekilde, son zamanlarda yapılan çalışmalar, test ve kişilik özelliklerinin kişi uyumu ile nasıl ilişkili olduğuna odaklanmıştır. Yakın tarihli bir çalışma (Dodeen ve Darabi, 2009), kişilik özelliklerinin normal olmayan yanıtlarla ilişkili olduğunu belirtmiştir. Örneğin, belirli kişilik özelliklerine sahip bir yanıt verici, bir kişiliği ölçen bir ölçüm aracına yanıt verirken daha iyi bir izlenim vermeye çalışabilir ve buna göre daha yüksek veya daha düşük puanlı yanıt seçeneklerini tercih

edebilir (Zickar ve Drasgow, 1996). Yakın zamanda yapılan bir başka çalışma, uyumsuz olmakla aykırı değer arasında bir ilişki olup olmadığı sorusunu ele almıştır (Felt, Castaneda, Tiemensma ve Depaoli, 2017) ve uyumsuz katılımcıların muhtemelen uç değere sahip olduğu sonucuna varmıştır. Buna ek olarak, araştırmacılar, kişi uyum istatistiklerinin uç değerleri tespit etmek için etkili bir teşhis aracı olabileceğini belirtti. Bu nedenle, yanıtlayıcıların uyumsuz yanıt kalıpları üretmelerinin bir başka olası nedeni, uç yanıt seçeneklerini kullanmalarıdır. Sonuç olarak, ilgili alanyazın, uygun olmayan yanıtlayıcıların uç yanıtları daha fazla kullanıyor olabileceğini desteklemektedir.

İlgili alan yazında, bireyin test maddelerine, maddelerin ölçmeyi amaçladığı özellikten farklı bir temelde belirli ve sistematik bir şekilde yanıt verme eğilimine uç tepki stili denmektedir (Paulhus 1991). Weijters (2006) yanıt stili kaynaklarını iki ana kategoriye ayırmıştır: uyaran düzeyi ve yanıtlayıcı düzeyi. Uyaran düzeyinde, yanıt stili ölçme aracının bir sonucu olarak görülürken, yanıtlayıcı düzeyinde yanıt stili kişisel özelliklerin bir sonucu olarak görülmektedir. Uyaran düzeyindeki olası yanıt stili kaynaklarına örnek olarak madde formatı (Weijters, Cabooter ve Schillewaert, 2010) ve verilerin toplandığı mod verilebilir. Katılımcı düzeyinde ise bireyin psikolojik ve demografik özellikleri temel kaynaklardır.

Yanıt stilleri genel olarak sistematik hataya sebep olurlar. Toplanan verilerin tek değişkenli ve çok değişkenli dağılımları üzerinde etkileri vardır. Bu da geçUYSiz ve hatalı yanıt örüntülerine neden olur. Bununla birlikte, yanıt stillerinin bu önemine rağmen, araştırmacılar sıklıkla çalışmalarında yanıt stillerinin incelenmesini analiz süreçlerine dâhil etmemektedir. Yanıt stilleri hakkında daha fazla bilgi edinmek isteyen okuyucuların ilgili alanyazına bakmaları önerilir (örn. Baumgartner ve Steenkamp, 2001; Hurley, 1998).

Yanıt stilleri içerisinde en yaygın olarak bilineni UYS'dir. Maddelerin içeriğinden bağımsız olarak Likert tipi veya derecelendirme ölçeklerinde uç noktaların orantısız bir şekilde tercihine karşılık gelmektedir. UYS'yi olası yanıt kategorileri arasında orta seçenekleri onaylamanın tUYSi olarak düşünmek de mümkündür. Örneğin, beşli Likert tipi bir ölçek maddesi için UYS, bir katılımcının “kesinlikle katılıyorum” veya “kesinlikle katılmıyorum” seçeneklerini tercih ettiği madde sayısı ile ilgilidir (Greenleaf, 1992). Bir yanıt verenin bir dizi maddedeki uç yanıtlarını basitçe toplamak UYS kullanımının miktarını hesaplamının geleneksel olarak kullanılan yöntemidir (Baumgartner ve Steenkamp 2001). Bu teknik “Sayma tekniği” olarak da bilinir. Vaerenbergh ve Thomas (2013) UYS kullanımı belirlemek için kullanılacak alternatif tekniklerin ayrıntılı bir incelemesini vermiştir.

UYS'yi kullanan kişilerin, yanıt verenin özellik düzeyleriyle kestirilemeyen yanıt örüntüleri üretmesi olasıdır. Ek olarak, uç yanıtların orantısız kullanımı, değişkenler arasındaki korelasyonları değiştirir ve bireysel farklılıkların doğasını anlamayı zorlaştırır. İç tutarlılığın hesaplanması, regresyon analizi, faktör analizi, yapısal eşitlik modelleri gibi birçok farklı analiz korelasyonlara dayanmaktadır (Baumgartner ve Steenkamp, 2001). Bu nedenle uç yanıtların etkisinin göz ardı edilmesi yanıltıcı sonuçlara yol açabilir. Önemi ve ölçüm sonuçlarının etkisine rağmen, gerçek dünya test uygulamalarında kişi uyum kontrolü sıklıkla başvurulan bir yöntem değildir (Meijer, Niessen ve Tendeiro, 2016).

Ayrıca, UYS'nin kullanımı ile kişilik arasındaki bağlantı ilgili alanyazında ima edilmiştir. Birçok çalışma, bireylerin UYS kullanımının zaman içinde nispeten sabit kaldığını göstermiştir (Jain ve Agrawal, 1977; Greenleaf, 1992). Ek olarak, önceki çalışmalar, öz bildirim ölçüm araçları kullanılırken UYS'nin kullanımının oldukça yaygın olduğunu ortaya koymuştur. Yakın tarihli bir araştırma, katılımcıların %25'inin yanıt süreci sırasında UYS'yi benimsediğini göstermiştir (Austin, Dear ve Egan, 2006). Bu yanıt stillerinin tutarlılığı ve istikrarı (ör. Weijters, Geuens ve Schillewaert, 2010), bireylerin özelliklerinin ve hatta kişiliklerinin UYS'yi kullanma üzerinde bir etkisi olabileceğini ima etmektedir. Daha önceki bulgulara göre, vicdanlılık ile uç yanıt stilleri arasında pozitif bir ilişki vardır. Ayrıca dışadönüklüğün UYS ile ilişkili olduğu bulunmuştur (Austin vd., 2006; Naemi, Beal ve Payne, 2009). Bulgular ayrıca He ve van de Vijver (2013) tarafından da doğrulanmıştır. Son olarak Hamilton (1968) tarafından yapılan bir başka çalışmada da dürtüsellik ile UYS arasında negatif bir ilişki olduğu belirtilmiştir (He ve van de Vijver, 2013). Sonuç olarak, kişilik ve UYS arasındaki ilişki konusunda tutarlı sonuçlara ulaşılmıştır (Kyllonen, 2008; Landers, Sackett ve Tuzinski, 2011).

Daha önce belirtildiği gibi, yanıt örüntülerini değerlendirmek, ölçüm hatasının kaynağını belirlemek için kritik öneme sahiptir ve Madde Tepki Kuramı (MTK) model tabanlı bir ölçüm yaklaşımı olduğundan, uyumsuz bireylerin saptanması MTK çerçevesi bağlamında özellikle önemlidir. Ayrıca, daha önce gerçekleştirilmiş çalışmalar, başarı sınavlarını değerlendirirken sıklıkla kişi uyumunu araştırmıştır (Brown ve Villareal, 2007). Ancak, psikometri alanındaki diğer yaklaşımlarda olduğu gibi, kişi uyum istatistikleri de ölçüm araçlarının çok yanıt kategorisine sahip maddeleri içerdiği psikolojik değerlendirme süreçlerinde kullanılmaya başlanmıştır (Dodeen ve Darabi, 2009; Ferrando, 2012). UYS, kullanılan ölçüm modeliyle uyumlu olmayan yanıt örüntüleri üreteceğinden, kişi uyum istatistikleri üzerinde olumsuz bir etkiye sahip olabilir. Bu olasılık doğrultusunda, bu

çalışmanın amacı, kişi-uyumu ile UYS arasında bir ilişki olup olmadığını araştırmak olarak belirlenmiştir.

Yöntem

Kişi uyumu ve uç tepki stili arasındaki ilişkiyi incelemeyi amaçlayan bu araştırma, bu iki değişken arasındaki bağlantıyı belirlemeyi amaçladığından ilişkiyel bir araştırmadır.

Katılımcılar

2020 küresel salgını nedeniyle araştırmaya katılım çevrimiçi olarak gerçekleştirilmiştir. Çevrimiçi veri toplama yaklaşımı kullanılırken sistematik örnekleme işlemlerinin uygulanması kolay olmadığından, veriler uygun örnekleme kullanılarak toplanmıştır. Katılımcı grubu büyük bir ilde bulunan dört farklı devlet üniversitesinden seçilmiş ve 11 farklı bölümde öğrenim gören 727 öğrenciden oluşmuştur. Örneklem grubunun demografik özellikleri aşağıdaki Tablo 1'de verilmiştir.

Buna göre, katılımcılar yaşları 16 ile 23 arasında değişen 590 kadın (Ort=21.19, SS=1.70) ve 17 ile 23 arasında değişen 137 erkek (Ort.=21.39, SS=1.47) oluşturmuştur. Bölümlere göre ise Tarih (n=32), Psikoloji (n=120), Diş Hekimliği (n=104), Çocuk Gelişimi (n=69), Psikolojik Danışmanlık (n=65), Fen Bilgisi Öğretmenliği (n=25), Matematik Öğretmenliği (n=102), Sosyal Bilgiler Öğretmenliği (n=42), Edebiyat (n=69), Sosyoloji (n=8), Güzel Sanatlar (n=97) öğrencilerinden oluşmaktadır. Bu nihai örneklem büyüklüğü, 73 katılımcının eksik veriye sahip olması ve uç değer olması nedeniyle elenmesiyle elde edilmiştir. Bu 73 katılımcının çıkarılmasının örneklem grubunun yaş, cinsiyet ve kayıtlı oldukları bölüm özelliklerini bozup bozmadığı incelenmiş ve çıkartılmalarının dağılımı etkilemediği görülmüştür.

Tablo 1. Katılımcıların demografik özelliklerine ilişkin tanımlayıcı istatistikler

Değişken	Gruplar	n	%
Cinsiyet	Kadın	590	81.2
	Erkek	137	18.8
Bölüm	Tarih	26	3.6
	Psikoloji	120	16.5
	Diş Hekimliği	104	14.3
	Çocuk Gelişimi	69	9.5
	Psikolojik Danışmanlık	65	8.9

Fen Bilgisi Öğretmenliği	25	3.4
Matematik Öğretmenliği	102	14.0
Sosyal Bilimler Öğretmenliği	42	5.8
Türk Dili ve Edebiyatı	69	9.5
Sosyoloji	8	1.1
Güzel Sanatlar	97	13.3

Veri Toplama Araçları

Giriş bölümünde sunulan alanyazın bir bütün olarak yeniden ele alındığında, yanıt verenler bir kişilik envanterine yanıt verdiklerinde, onların muhtemelen UYS'yi kullanacaklarını veya uyum gösteremeyen katılımcı yanıtlar üreteceklerini varsaymak doğru olacaktır. Bu nedenle, UYS ile kişi uyumu arasındaki ilişkiyi araştırmak için bir kişilik envanteri ile veri toplanması tercih edilmiştir. Buna doğrultuda veri toplama aracı olarak Temel Kişilik Özellikleri Envanteri (TKÖE) seçilmiştir. TKÖE bir öz bildirim aracıdır ve Gençöz ve Öncül (2012) tarafından üniversite Öğrencilerinin kişilik özelliklerini belirlemek amacıyla geliştirilmiştir. Kırk beş sıfat maddesi ile ölçülen ve altı boyuttan oluşan beşli Likert türünde puanlanan maddelere sahip bir ölçektir. Envanterde katılımcılara bu sıfatların kendilerini ne ölçüde temsil ettiği sorulmuştur. Ölçek, beş boyutu (Dışa Dönüklük-D, Vicdanlılık-V, Uyumluluk-U, Nörotisizm-N ve Açıklık-A) içeren geleneksel beş kişilik boyutunun yanı sıra altıncı bir boyut daha içerir: Olumsuz Değerlilik-OD. Geliştirme aşamasında, TKÖE'nin faktör yapısını incelemek amacıyla açımlayıcı faktör analizi yapılmış ve sonuçlar altı boyutun toplam varyansın %53'ünü açıkladığını ve maddelerin aldığı yüklerin .84 ile .40 arasında değiştiğini ortaya koymuştur. KKÖE, Gençöz ve Öncül tarafından iç tutarlılık ve test-tekrar test güvenilirliği açısından da değerlendirilmiş ve alt boyutlar için iç tutarlılık katsayılarının .71 ile .89 arasında değiştiği bulunmuştur. Test-tekrar test güvenilirlik değerleri de .71 ile .84 arasında değişmektedir. Bu katsayı değerleri, TKÖE maddeleri için kabul edilebilir güvenilirlik düzeyini desteklemiştir. Mevcut çalışmada, güvenilirlik kavramına farklı bir şekilde yaklaşan MTK bağlamında kişi-uyumu araştırıldığından, geleneksel yaklaşımın aksine TKÖE maddelerinin iç tutarlılığı yeniden hesaplanmamıştır.

Veri Analizi

Kişi-uyum alanyazını incelendiğinde bugüne kadar geliştirilmiş birçok farklı istatistiğin olduğu görülmüştür. İlgili okuyucuların bu yöntemleri gözden geçirmek için Meijer ve

Sijtsma'ya (2001) bakmaları önerilir. Mevcut çalışmada, en çok çalışılan ve en çok bilinen olduğu için lz kişi-uyum istatistiği (Drasgow, Levine ve Williams, 1985) tercih edilmiştir. Ayrıca, bu istatistik madde yanıt kuramı olarak da bilinen güçlü bir psikometrik yaklaşıma dayanmaktadır. Veri seti elde edildikten sonra, her katılımcı için lz istatistikleri hesaplanmıştır.

lz istatistiği, bireylerin yanıt vektörlerinin standartlaştırılmış log-olasılığıdır ve standartlaştırılmış z değerleri ile benzer bir dağılıma sahiptir ve yüksek negatif değerler uyumsuzluk olarak yorumlanır. lz istatistikleri hesaplanırken, bir kişinin uyum gösteren mi yoksa göstermeyen mi olduğuna dair bir kararın verildiği bir eşik değeri belirlemek gerekir. Mevcut çalışma için, eşik değerini belirlemek için $\alpha = 0.05$ düzeyi ile bootstrap yöntemi kullanılmıştır.

lz istatistikleri hesaplanıp eşik değerleri belirlendikten sonra, 0'ın uyum göstermeyen ve 1'in uyum gösteren kişileri temsil ettiği lz değerlerine dayalı yeni bir iki kategorili değişken oluşturulmuştur. Ek olarak, her bir birey için alt UYS (kişinin en düşük yanıt kategorisini seçip seçmediğini gösteren) ve üst UYS (kişinin en yüksek yanıt kategorisini seçip seçmediğini gösteren) kullanımını belirtmek için iki tane iki kategorili değişken daha oluşturulmuştur. Bu işlem ve TKÖE'nin her bir maddesi için tekrarlanmıştır. Bir kişi alt UYS veya üst UYS gösteriyorsa, bu kişinin genel olarak UYS kullandığı kabul edilmiştir. Son olarak, TKÖE'nin her maddesi için, iki kategorili UYS değişkeni ve iki kategorili kişi-uyum değişkenleri arasında ki-kare istatistikleri hesaplanmıştır. Bu şekilde UYS ile kişi-uyum arasındaki ilişkinin görülmesi amaçlanmıştır. lz istatistiklerinin hesaplanması, R istatistik programında (v4.0.0; R Core Team, 2019) bulunan "PerFit" paketi (v0.3.4; Tendeiro, Meijer ve Niessen, 2016) kullanılarak gerçekleştirilmiştir. Diğer istatistiksel işlemler için ise SPSS, sürüm 21 kullanıldı.

Bulgular

TKÖE'nin boyutları arasındaki Pearson korelasyon katsayı değerleri, her bir boyuta ait puanların ortalama, standart sapma, çarpıklık, basıklık değerleri aşağıdaki Tablo 2'de verilmiştir. Tüm boyutlar arasındaki korelasyon değerleri anlamlı bulundu ($p < 0.01$). Ayrıca basıklık ve çarpıklık değerleri incelendiğinde bu değerlerin dağılımlarının +1 ve -1 aralığında olduğu görülmüştür ki bu da alt boyut puanlarının dağılımının normal kabul edilebileceği anlamına gelmektedir.

Tablo 2. Kişilik Tipleri Arasındaki Korelasyonlar ve Betimleyici İstatistikler

	Boyutlar	1	2	3	4	5	6
1	Dışa dönüklük		0.28**	0.30**	-0.21**	0.58**	-0.22**
2	Vicdanlılık			0.32**	-0.22**	0.40**	-0.23**
3	Uyumluluk				-0.17**	0.38**	-0.40**
4	Nörotisizm					-0.17**	0.34**
5	Açıklık						-0.15**
6	Olumsuz değerlilik						
	Ort=	18.72	22.35	33.57	27.81	29.39	34.15
	SS=	4.92	6.06	8.34	6.45	5.73	4.14
	Çarpıklık=	-0.10	-0.38	-0.16	-0.20	-0.54	-0.54
	Basıklık=	-0.01	-0.40	-0.19	-0.56	0.09	0.07

** $p < 0.01$

İz değerlerinin betimleyici istatistikleri, bir katılımcının uyum gösterip göstermediğine karar vermek için belirlenen kesme noktası değerleri ve her bir alt boyut için uyum gösteren ve uyum göstermeyen olarak kodlanan katılımcı sayısını gösteren frekans istatistiklerine aşağıdaki Tablo 3'te verilmiştir. En fazla uyum göstermeyen kişi sayısı vicdanlılık (n=66) alt boyutunda, en az uyum göstermeyen kişi sayısı (n=49) olumsuz değerlilik alt boyutunda gözlenmiştir. Ayrıca boyutlar arasındaki korelasyon ile uyum göstermeyen kişi sayısının benzerliği arasındaki bağlantı niteliksel olarak değerlendirilmiştir. Bu ön inceleme nihai sonucun bir elde edilebileceğini göstermektedir.

Tablo 3. *Kestirilen Kişi Uyum İstatistikleri İçin Betimleyici İstatistikler*

	Min.	Maks.	Ort.	SS	Çarpıklık	Basıklık	Kesim	Uymlu	Uymsz
Dışa dönüklük	-4.74	1.89	0.27	1.10	-1.26	2.12	-1.48	670	57
Vicdanlılık	-7.31	1.97	0.25	1.19	-1.60	3.56	-1.42	671	66
Uyumluluk	-9.89	1.64	0.18	1.20	-2.54	12.20	-1.50	674	53
Nörotisizm	-5.55	2.25	0.30	1.10	-1.30	2.80	-1.36	667	60
Açıklık	-5.52	1.50	0.27	1.03	-1.52	2.80	-1.52	677	50
Olumsuz değerlilik	-3.62	1.38	0.10	0.75	-1.45	2.73	-1.12	678	49

UYS ile kişi-uyum arasındaki ilişkiyi incelemek amacıyla TKÖE'deki her bir madde için Ki-kare istatistiklerini hesaplamak için iki kategorili değişkenler (UYS kullanma vs. UYS kullanmama ve uyum gösterme vs uyum göstermeme) kullanılmıştır. UYS değişkeni, bir kişi en düşük yanıt kategorisini (alt UYS olarak adlandırılır) veya en yüksek yanıt kategorisini (üst UYS olarak adlandırılır) kullanıyorsa, her iki durum için de ayrı olarak ele alınmıştır. Sonuçlar aşağıdaki Tablo 4'te verilmiştir. Bulgulara dayalı olarak 24 madde için uyum göstermeme ve UYS kullanmanın bağımlı olduğu anlaşılmıştır. Başka bir deyişle, bu maddeler için uyum gösterme ya da göstermeme UYS'nin kullanılmasına önemli ölçüde bağlıdır.

Bu bulgu, uyum gösteren ya da göstermeyen katılımcılar arasında UYS kullanım sıklığının farkı gösterdiğini söylüyor. Tablo daha detaylı incelendiğinde 24 maddeden 17'sinde uyum göstermeyen bireylerin hem alt hem de üst UYS'yi uyum gösteren kişilere göre daha sık kullandıkları görülmüştür. Öte yandan, uyum gösteren katılımcılar için 5 maddede alt UYS kullanılırken, sadece 2 madde için üst UYS daha fazla kullanılmıştır. Kısacası UYS'nin uyum göstermeyen bireyler tarafından daha yoğun kullanıldığı söylenebilir.

Tablo 4. *Kişi Uyum Gruplarına Göre Yanıt Stillerinin Madde Düzeyinde Karşılaştırılması*

Maddeler	UYS (alt)		UYS (Üst)			p
	Uymsz	Uyuml	Uymsz	Uyuml	Uymsz	
Dışadönüklük						
Ürkek	10.50	6.10	4.40	17.50	14.00	3.50
Geri çekildi	68.40	70.40	-2.00	0.00	0.10	-0.10
Utangaç	0.00	0.60	-0.60	31.60	35.10	-3.50

Konuşkan	10.50	3.00	7.50	35.10	21.80	13.30	**
Uyuşuk	1.80	1.50	0.30	19.30	20.90	-1.60	
Girişimci	15.80	4.20	11.60	24.60	27.90	-3.30	**
Soğuk	17.50	8.50	9.00	19.30	11.30	8.00	*
Pasif	3.50	0.70	2.80	47.40	41.60	5.80	
Dürüstlük							
Öz disiplin	18.20	14.40	3.80	21.20	11.50	9.70	
Derli toplu	4.50	2.30	2.20	18.20	14.20	4.00	

Çalışkan	22.70	30.00	-7.30	15.20	3.80	11.40	**
İhtiyatlı	10.60	3.20	7.40	28.80	14.70	14.10	**
Telaşlı	60.60	66.00	-5.40	4.50	1.80	2.70	**
Azimli	9.10	4.70	4.40	19.70	14.80	4.90	
Sorumsuz	4.50	0.00	4.50	65.20	49.50	15.70	**
Tembel	1.50	0.20	1.30	68.20	47.70	20.50	**
Uyumluluk							
İçten	0.00	1.60	-1.60	32.10	26.10	6.00	
Şefkatli	13.20	20.80	-7.60	17.00	5.80	11.20	*
Güler yüzlü	3.80	0.00	3.80	45.30	45.80	-0.50	**
İyi niyetli	5.70	1.90	3.80	37.70	21.50	16.20	**
Hayırsever	15.10	23.10	-8.00	9.40	8.30	1.10	
Hoşgörülü	9.40	2.40	7.00	50.90	46.30	4.60	*
Paylaşan	7.50	0.60	6.90	37.70	37.20	0.50	**
Duyarlı	3.80	1.20	2.60	34.00	38.40	-4.40	
Nevrotiklik							
Gergin	10.00	4.50	5.50	21.70	21.10	0.60	
Agresif	61.70	62.40	-0.70	3.30	1.30	2.00	
Sinirli	31.70	15.10	16.60	35.00	7.80	27.20	**
Mizaçlı	1.70	0.30	1.40	43.30	34.80	8.50	
Sabırsız	35.00	16.80	18.20	25.00	5.80	19.20	**
Kaprisli	48.30	31.90	16.40	11.70	3.60	8.10	**
Aceleci	80.00	81.70	-1.70	0.00	0.10	-0.10	
Dokunaklı	6.70	4.60	2.10	35.00	16.80	18.20	**

Endişeli	23.30	11.50	11.80	16.70	10.50	6.20	*
Açıklık							
Kendine güvenen	86.00	80.80	5.20	4.00	0.60	3.40	*
Kendinden emin	8.00	13.90	-5.90	30.00	16.20	13.80	*
Cesur	12.00	2.40	9.60	28.00	18.20	9.80	**
Yaratıcı	36.00	37.50	-1.50	12.00	3.40	8.60	*
Kolay ve kesintisiz							
ilerleme	12.00	5.80	6.20	22.00	27.00	-5.00	
Yetenekli	6.00	6.20	-0.20	30.00	22.30	7.70	
Olumsuz değerlilik							
Terbiyesiz	12.00	17.70	-5.70	14.00	9.70	4.30	
İddialı	2.00	0.40	1.60	42.00	39.70	2.30	**
Kaba	6.00	3.40	2.60	40.00	28.80	11.20	
Arkadan bıçaklama	8.00	3.10	4.90	36.00	27.90	8.10	
Aç gözlü	4.00	1.80	2.20	42.00	30.10	11.90	
Gizlenmiş	8.00	1.50	6.50	42.00	30.70	11.30	**

Not: * $p < 0,05$; ** $p < 0,01$; UYS (düşük)=en yüksek en düşük kategoriye onaylayan katılımcıların yüzdesi; UYS (Üst)= en yüksek kategoriye onaylayan katılımcıların yüzdesi; Diff.= uygun ve uyumsuz grup arasındaki yüzde değerleri farkı; Tablodaki tüm değerler yüzdeleri gösterir.

Kişi-uyum ve UYS'nin kullanımı arasındaki ilişkiyi açıkça belirtmek için, en yüksek ve en düşük uyuma sahip kişiler (en düşük ve en yüksek lz fit istatistik değerlerine sahip olanlar) için yanıt matrislerinin küçük bir kısmı Tablo 5'te sunulmuştur. Tabloda ayrıca her iki grup tarafından ortalama üst ve alt UYS kullanımına yer verilmiştir. Sonuçlar, en yüksek uyum göstermeme düzeyine sahip 10 kişinin bu 8 maddelik örnekte ortalama 1.7 defa alt UYS kullandığını, üst UYS kullanım ortalamasının ise 2.6 olduğunu göstermektedir. Uyum düzeyi en yüksek olan katılımcılar için ise alt UYS kullanım ortalaması 0.9 iken bu ortalama üst UYS kullanımı için 1.1'dir. Yani uyum göstermeyen kişiler, alt ve üst UYS kullanımı açısından daha fazla ayrılmaktadır. Bu kişiler genellikle üst UYS'yi daha sık kullanırlar. Ayrıca, Ki-kare testi, uyum gösterip göstermemenin üst veya alt UYS kullanımı ile ilgili olup olmadığını test etmek için kullanılmıştır. Sonuç, bu ilişki için kanıt olmadığını göstermektedir ($p > 0,05$). Yani, bir yanıtlayıcı hangi tür UYS kullanırsa kullansın, uyum göstermeyen kişiler uyum gösterenlerden daha fazla UYS kullanır. Bu bulgu, en düşük lz

değerine sahip bireylerde (en yüksek düzeyde uyum göstermeyen katılımcılar) UYS kullanımının daha yoğun olduğunu göstermiştir. Bu bulgu hem üst hem de alt UYS kullanımı için geçerlidir.

Tablo 5. En Yüksek ve En Düşük İz değerine sahip kişiler İçin Ortalama UYS Kullanımı ve Yanıt Matrisleri

	i1	i2	i3	i4	i5	i6	i7	i8	UYS (alt) #	UYS (üst) #
Uymsz Kişi 1	5	5	5	1	1	5	5	5	2	6
Uymsz Kişi 2	5	5	5	4	5	5	5	5	0	7
Uymsz Kişi 3	1	3	3	3	1	3	5	5	2	2
Uymsz Kişi 4	5	5	3	1	1	3	5	5	2	4
Uymsz Kişi 5	1	4	3	3	1	4	1	1	4	0
Uymsz Kişi 6	2	2	4	2	2	1	2	4	1	0
Uymsz Kişi 7	5	4	3	5	1	2	4	5	1	5
Uymsz Kişi 8	4	3	1	4	4	4	4	4	1	0
Uymsz Kişi 9	3	4	1	2	1	4	5	5	2	2
Uymsz Kişi 10	3	4	4	1	1	4	2	3	2	0
									Ort. 1.7	2.6
Uymlu Kişi 1	3	4	4	4	1	4	5	5	1	2
Uymlu Kişi 2	3	5	3	4	1	4	5	5	1	3
Uymlu Kişi 3	4	2	2	4	2	3	4	4	0	2
Uymlu Kişi 4	4	5	3	4	3	4	5	4	0	2
Uymlu Kişi 5	2	4	1	4	1	4	4	4	2	0
Uymlu Kişi 6	3	4	4	4	2	3	4	4	0	0
Uymlu Kişi 7	2	3	3	4	1	3	4	4	1	0
Uymlu Kişi 8	4	5	3	4	1	5	4	5	1	2
Uymlu Kişi 9	2	3	3	3	1	3	3	3	1	0
Uymlu Kişi 10	1	4	3	3	1	2	3	3	2	0
									Ort. 0.9	1.1

Tartışma

Girişte bahsedildiği gibi, insanların tepki stilleri, sahip oldukları özellik düzeyinin göstergesi olmayan puanlar almalarına neden olabilir. Madde tepki kuramı, ölçüm için modele dayalı

bir yaklaşım kullandığından, yanıt kalıpları diğer süreçlerden etkilenirse, kişilerin kestirilen yetenek düzeylerinin modele göstermez hale gelmesi muhtemeldir. Bu durumda bu bireyler belirtilen modele uymamaktadır. Bu çalışmada, tepki stiline kişi uyumu ile nasıl bir ilişkisi olduğu araştırılmıştır. Buna göre bireylerin uyum düzeyini gösteren IZ istatistikleri hesaplanarak iki kategoride (uyum ve uyumsuzluk) yeniden kodlanmıştır. Ayrıca bireylerin tepki örüntüleri incelenmiş ve her bir madde için alt veya üst UYS'nin kullanıldığını belirten değişkenler oluşturulmuştur. Daha sonra bu kategorik değişkenler arasındaki ilişkiler her bir madde için ki-kare istatistiği ile incelenmiştir. Sonuç olarak, maddelerin çoğu için, kişi-uyum ve UYS arasında istatistiksel olarak anlamlı bir ilişki bulundu. Benzer bir araştırma sorusu bir simülasyon çalışmasında araştırılmış ve benzer ve karşılaştırılabilir sonuçlar elde edilmiştir (Tendeiro, 2017; Adams vd., 2019). Bu önceki çalışmadan farklı olarak, mevcut çalışma gerçek verileri kullanılmış ve kişi uyumu ile UYS arasındaki bağlantı olduğu daha da desteklenmiştir.

Ayrıca en yüksek ve en düşük IZ değerlerine sahip 10 kişinin yanıt örüntüleri incelenmiş ve UYS kullanımlarına göre karşılaştırılmıştır. Benzer şekilde uyumsuz kişilerin daha fazla UYS kullandığı doğrulanmıştır. Bulgular daha detaylı incelendiğinde en tutarlı bulguların nevroitiklik alt boyutundaki maddelerde gözlemlendiği söylenebilir. Bu boyut için UYS ile kişi uyum arasındaki ilişki incelendiğinde, bu boyut için tüm maddelerde hem alt UYS hem de üst UYS'nin uyum göstermeyen kişilerde daha fazla gözlemlendiği görülmüştür. Bu temel sonuç, uç değerler ve kişi uyumu arasındaki bağlantının araştırıldığı önceki alanyazınla (bkz. Felt ve diğerleri, 2017) uyumludur. Uç değerler mutlaka aşırı tepkiler kullanmayabilir fakat uç değer olmak ve uç yanıt kullanmak oldukça örtüşen kavramlardır.

Sonuçlar

Bu çalışmada, bireylerin normal olmayan tepkilerini belirlemek için tek bir ölçme aracı ve uyum göstermeyen kişileri belirlemek için sadece bir kişi uyum istatistiği kullanıldığı için sonuçlar dikkatle okunmalıdır. Yalnızca bir kişi istatistik kullanmanın olumsuz etkilerini daha iyi görmek için okuyucuların St-Onge ve arkadaşlarına (2011) bakmaları önerilir. Giriş bölümünde belirtildiği gibi alan yazında birçok farklı uyum istatistiği bulunmaktadır. Ancak bu alternatif uyum istatistiklerini ve farklı kişilik ölçeklerini aynı çalışmada kullanmak, elde edilen sonuçların geçerliliğini artıracaktır.

Geçmişte bu konuda nispeten az sayıda çalışma bulunduğundan, bu çalışma diğer ön çalışmalar temelinde yürütülmemiş ve tartışılmamıştır. Sonuç olarak okuyucuların bulgulara yönelik alternatif açıklamalar ve yorumlar geliştirmeleri mümkündür. Bu çalışmada ulaşılan

sonuçların genişletilmesi gerekmektedir. Mevcut çalışmada elde edilen bulguların bundan sonraki çalışmalarla doğrulanması veya çürütülmesi gerektiği söylenebilir. Ek olarak, bu çalışmada mevcut yanıt stilleri arasında sadece UYS dikkate alınmıştır. Giriş bölümünde belirtildiği gibi, birçok farklı yanıt stili vardır ve bu yanıt stillerinin kişi uyumu ile ilgili her birinin incelenmesi mevcut bulgulara katkıda bulunacaktır. Son olarak, gerçek verilerle çalışmanın bir sonucu olarak, katılımcıların yanıtlarını etkileyen tüm süreçler kontrol edilememiştir. Bu nedenle, gözlemlenen yanıt örüntülerini UYS bağlamında değerlendirmek yanıltıcı sonuçlar sağlayabilir. Örneğin, UYS kullanımının dikkatsiz tepkilerden mi yoksa katılımcıların kendilerini sosyal olarak daha çekici gösterme arzusundan mı kaynaklandığı kontrol edilmemiştir. Bu nedenle sonuçların dikkatli bir şekilde yorumlanması önerilir.

Bu çalışmada kişilik verileri kullanılmış olmasına rağmen, psikolojik değişkenlerin eğitimsel çıktılarıyla yakından ilişkili olması ve kişilik ile akademik başarı arasındaki ilişkinin alan yazında hâlihazırda raporlanmış olması nedeniyle sonuçlar sınırlı düzeyde de olsa pedagojik çıkarımlar yapmamızı sağlayabilir (Nambudiri, Shaik ve Gülyani, 2019). Veri setinden uyum göstermeyen bireylerin çıkarılmamasının hatalı teşhislere yol açabileceği ve yanıltıcı sonuçların eğitimcilerin öğrencilerin eğitimsel çıktılarını iyileştirmek için proaktif adımlar atmasını engelleyebileceği unutulmamalıdır. Psikolojik değişkenlerin geçerli ve güvenilir ölçümü olmadan eğitim sonuçlarının iyileştirilemeyeceği açıktır.

Karabatsos (2003), %25 ve üzeri normal olmayan yanıt oranının kabul edilebilirlik açısından eşik olduğunu belirtmiştir. Bu çalışmada uyumsuz bireylerin oranı %10'un altında bulunmuştur. Bu nedenle, kişi uyum oranının kabul edilebilir eşik çok altında olduğu mevcut koşullarda, herhangi bir değişken bağlamında incelenmesi kesin sonuçlar çıkarmak için sağlıklı olmayabilir.

Sonraki çalışmalarda üçüncü bir değişkenin etkisi incelenebilir. Örneğin, UYS ile kişi uyumu arasındaki ilişkinin cinsiyet değişkenine göre nasıl değiştiği sorusunun incelenmesi bu konudaki mevcut bilinenleri zenginleştirecektir. Ayrıca her bir kişilik alt boyutu için elde edilen puanlara göre kişi uyumu ve UYS arasındaki ilişkinin nasıl değiştiğinin incelenmesi faydalı olacaktır. Bu sayede bireylerin kişilik özelliklerinin etkisi mevcut bulgularla ilişkilendirilebilir. Son olarak, UYS'nin kişi uyumuna göre farklılık gösterdiği maddelerin model uyum düzeylerinin incelenmesi önerilmektedir. UYS'nin MTK temelli madde uyumunu nasıl etkilediğini kişi uyumu bağlamında incelemenin mevcut alanyazına büyük katkı sağlayacağına inanılmaktadır.

Etik Kurul İzin Bilgisi: Bu araştırma Marmara Üniversitesi Eğitim Bilimleri Enstitüsü Araştırma ve Yayın etik kurulunun 19.11.2020 tarih ve 2020.8.1.82 sayılı kararı ile izni ile yapılmıştır.

Yazar Çıkar Çatışması Bilgisi: Yazar herhangi bir çıkar çatışması beyan etmemektedir.

Kaynakça

- Adams, D. J., Bolt, D. M., Deng, S., Smith, S. S., & Baker, T. B. (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of Mathematical and Statistical Psychology*, 72(3), 466-485.
- 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(6), 1235-1245.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 143-156.
- Brown, R. S., & Villarreal, J. C. (2007). Correcting for person misfit in aggregated score reporting. *International Journal of Testing*, 7(1), 1-25.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475-494.
- Cronbach, L. J. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement*, 10(1), 3-31.
- Dodeen, H., & Darabi, M. (2009). Person-fit: relationship with four personality tests in mathematics. *Research Papers in Education*, 24(1), 115-126.
- Drasgow, F., Levine, M. V., and Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38(1), 67-86.
- Embretson, S. E., Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates.

- Felt, J. M., Castaneda, R., Tiemensma, J., and Depaoli, S. (2017). Using Person Fit Statistics to Detect Outliers in Survey Research. *Frontiers in psychology*, 8, 863.
- Ferrando, P. J. (2007). Factor-analytic procedures for assessing response pattern scalability. *Multivariate Behavioral Research*, 42(3), 481-507.
- Ferrando, P. J. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. *Personality and Individual Differences*, 52(6), 718-722.
- Ferrando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement*, 61(6), 997-1012.
- Gençöz, T. & Öncül, Ö. (2012). Examination of Personality Characteristics in a Turkish Sample: Development of the Basic Personality Traits Inventory. *The Journal of General Psychology*, 139(3), 194-216.
- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56(3), 328-351.
- Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology*, 132(2), 201-210.
- International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. *International Journal of Testing*, 14(3), 195-217.
- Jain, U., & Agrawal, L. (1977). Generality of extreme response style. *Journal of psychological researches*, 21, 67-72.
- Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. M. (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software*, 74(1), 1-27.
- Tendeiro, J. N. (2017). The lz (p)* Person-fit statistic in an unfolding model context. *Applied Psychological Measurement*, 41(1), 44-59.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277-298.
- Kyllonen, P. C. (2008). *The research behind the ETS Personal Potential Index (PPI)*. Princeton, NJ: Educational Testing Service.

- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology, 96*(1), 202.
- Lomax, R. G. (1986). The effect of measurement error in structural equation modeling. *The Journal of Experimental Education, 54*(3), 157-162.
- Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*(2), 107-135.
- Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2016). A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program. *Assessment, 23*(1), 52-62.
- 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*(1), 93-119.
- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of personality, 77*(1), 261-286.
- Nambudiri, R., Shaik, R., & Ghulyani, S. (2019). Student personality and academic achievement: mediating role of psychological capital (PsyCap). *International Journal of Educational Management, 4*(34): 767-781.
- Paulhus, D. L. (1991), "Measurement and control of response bias," in *measures of personality and social psychological attitudes*, John P. Robinson, Phillip R. Shaver, and Lawrence S. Wright, eds. San Diego: Academic Press, 17-59.
- Petridou, A., & Williams, J. (2007). Accounting for aberrant test response patterns using multilevel models. *Journal of Educational Measurement, 44*(3), 227-247.
- Reise, S. P., and Waller, N. G. (1990). Applications of response pattern aberrancy analysis to personality assessment. *Unpublished manuscript*.
- Reynolds, C. R. (2010). Measurement and assessment: An editorial view. *Psychological Assessment, 22*, 1-4.

- Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, 55(1), 3.
- St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2011). Accuracy of person-fit statistics: A Monte Carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, 35(6), 419-432.
- Tatasuoka, K. (1996). Use of generalized person-fit indexes, zetas for statistical pattern classification. *Applied measurement in education*, 9(1), 65-75.
- Pan, T., & Yin, Y. (2017). Using the Bayes factors to evaluate person fit in the item response theory. *Applied Measurement in Education*, 30(3), 213-227.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International journal of public opinion research*, 25(2), 195-217.
- Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks: Sage Publications.
- Weijters, B. (2006). *Response styles in consumer research*. Doctoral Dissertation. Ghent University. Faculty of Economics and Business Administration, Ghent, Belgium.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236-247.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological methods*, 15(1), 96.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied psychological measurement*, 20(1), 71-87.